

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 20:53:41 ; Search time 53 Seconds
(without alignments)
8275.405 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattggaataca.....tttaacgacttacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PT05_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	103.2	6.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	76	4.5	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 3	71.2	4.2	152331	3 US-09-128-155-16	Sequence 16, Appl
C 4	71.2	4.2	176373	3 US-09-128-155-17	Sequence 17, Appl
C 5	62.6	3.7	152331	3 US-09-128-155-16	Sequence 16, Appl
C 6	58	3.4	6124	4 US-08-213-419B-3	Sequence 3, Appl
C 7	56.8	3.3	289	4 US-09-007-005-17	Sequence 17, Appl
C 8	56.8	3.3	289	4 US-09-244-796-17	Sequence 17, Appl
C 9	55.6	3.3	291	1 US-07-922-723A-7	Sequence 7, Appl
C 10	55.6	3.3	291	1 US-07-799-828C-7	Sequence 7, Appl
C 11	55.6	3.3	291	1 US-08-074-275-7	Sequence 7, Appl
C 12	55.6	3.3	291	1 US-08-480-366-7	Sequence 7, Appl
C 13	55.6	3.3	291	2 US-07-952-277A-7	Sequence 7, Appl
C 14	55.6	3.3	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 15	55.6	3.3	44453	3 US-09-146-053-5	Sequence 5, Appl
C 16	54.2	3.2	5852	1 US-07-867-106-2	Sequence 2, Appl
C 17	53.4	3.1	1956	4 US-08-559-896B-1	Sequence 1, Appl
C 18	52.8	3.1	7379	4 US-09-341-587-5	Sequence 5, Appl
C 19	52.4	3.1	1298	4 US-08-971-089-1	Sequence 1, Appl
C 20	51.8	3.0	454	2 US-08-623-906A-6	Sequence 6, Appl
C 21	51.6	3.0	72604	4 US-09-268-992-7	Sequence 7, Appl
C 22	51.6	3.0	72604	4 US-09-657-474-7	Sequence 7, Appl
C 23	50.4	3.0	1559	4 US-09-019-095A-7	Sequence 7, Appl
C 24	50.2	3.0	2791	4 US-09-570-367C-1	Sequence 1, Appl
C 25	50	2.9	1298	3 US-08-948-705-3	Sequence 3, Appl
C 26	49.6	2.9	417	4 US-08-559-896B-3	Sequence 3, Appl
C 27	49.6	2.9	1859	3 US-08-691-563C-46	Sequence 46, Appl

28	49.6	2.9	45546	4 US-09-146-053-6	Sequence 6, Appl
29	49.2	2.9	2223	1 US-08-257-073-4	Sequence 4, Appl
30	49	2.9	2040	1 US-08-599-252-103	Sequence 103, App
31	49	2.9	2040	5 PCT-US96-06352-103	Sequence 103, App
32	49	2.9	2040	5 PCT-US96-06583-103	Sequence 103, App
33	48.6	2.9	12141	4 US-09-488-671-10	Sequence 10, Appl
C 34	48.2	2.8	372	4 US-09-018-584A-13	Sequence 13, Appl
C 35	47.8	2.8	658	4 US-08-998-416-595	Sequence 595, App
C 36	47.8	2.8	72604	4 US-09-268-992-7	Sequence 7, Appl
C 37	47.8	2.8	72604	4 US-09-657-474-7	Sequence 7, Appl
C 38	47.2	2.8	2447	2 US-09-014-969-14	Sequence 14, Appl
C 39	46.8	2.8	2030	4 US-09-512-342-1	Sequence 1, Appl
C 40	46.6	2.7	240	1 US-08-628-417-6	Sequence 6, Appl
C 41	46.6	2.7	1214	4 US-08-817-913-13	Sequence 13, Appl
C 42	46.6	2.7	1232	4 US-08-817-913-14	Sequence 14, Appl
C 43	46.6	2.7	1352	4 US-08-817-913-15	Sequence 15, Appl
C 44	46.6	2.7	1734	4 US-08-817-913-16	Sequence 16, Appl
C 45	46.6	2.7	1920	4 US-08-817-913-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzpgt-fls
; US-08-232-463-14

Query Match 6.1%; Score 103.2; DB 1; Length 7218;


```
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match      3.4%; Score 58; DB 4; Length 6124;
Best Local Similarity 46.9%; Pred. No. 0.00044;
Matches 215; Conservative 0; Mismatches 240; Indels 3; Gaps 1;

QY   107 AAATAATGTCATTTTAAATCAGACATTTGTTTTAAATCAAACTTAATCTCTTATATC 166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2169 AATTAATAAATTTAATAATAATAATAATATATATTTATTTAAATTTATATATATATATA 2228

QY   167 ACAACGACATTCGCGAAAATTCAGGTAAAAAGAGAAAATAAAGAATGAGAGATAGAGAG 226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2229 ATAAATTTTTTATTTTAAATTTAAATTAGATTGTCCTCCAAAAAATAAATAAATAAAT 2288

QY   227 ATTTCATGAAAAAGAGAGAGAACATGTAGTGGAACAATAAAGAGATATGATCAT 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2289 ATATATATATATTAATAATACATATATATTAACATAAGAAAAATTAATAATAATCAAAC 2348

QY   287 ATATTTATGAGAGGTGGTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAA 346
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2349 ATATTCAAAAAATAAAGCTTCTTAAATATATATATAACTTTAATATTCATATATCAAAAT 2408

QY   347 AATGACATGGTGAATCTGAAGAGATGAATTCGTTAAAGATGAGAGAGAAAGAGAACT 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2409 GAAGTCATATATTTCTTCTGTTTCATATGTGTAAGAAATGAAAAAATAAATAAAGGA 2468

QY   407 CCATGCTAAAGTCTCGTAAAGAAGATGAAAAAGAAACAAAGAAGAGGAAAGAGAGA 466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2469 AAAGGAAAAAGAAAAAGGAAAAAGAAAAAGAAAAAGAAAAACAAATATGTAAAAAT 2528

QY   467 AAGGCTAAATAGACTAATATGCAAAATTCGTAGCCGACAAATATCTATTTGGTCC 526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2529 ATAATTATTATATAATAATAAT--ATAATAATTTTACGCATACACAAACATTGTGCAT 2585

QY   527 AAGGTTATTTGTGTATCTTTTGAAGTCAAAAGTTAT 564
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2586 TATTTTTTTTAGTGTTATATTAAACAAAATGTTAT 2623

RESULT 7
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      3.3%; Score 56.8; DB 4; Length 289;
Best Local Similarity 12.2%; Pred. No. 0.00035;
Matches 31; Conservative 105; Mismatches 118; Indels 0; Gaps 0;

QY   195 AAAAGAGAAAAATAAAGAATGAGAGATAGAGAGATTTCTTATGAAAAAGAGAGAGAAC 254
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    33 RARCARARURURARCARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 92

QY   255 TGTAGTGTAACAAAATAAAGAGATATGATATATTTTATGAGAGGTGTGGAAGATTAT 314
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    93 RNPNRSRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSFNRNRSFNRSRNRNRS 152

QY   315 TTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAAGATGA 374
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    153 RNPNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSFNRNRSFNRSRNRNRS 212

QY   375 ATTGTTTAAAGATCAAGAGAGAAAGAACTCCATGCTGCTAAAGTCTCTGTAAGAAGATG 434
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    213 RNPNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSFNRNRSFNRSRNRNRS 272

QY   435 AAAAAGAAAAACAAA 448
      ||||| ||| ||| |||
Db    273 AAAAAAAAAAAAAA 286

RESULT 8
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      3.3%; Score 56.8; DB 4; Length 289;
Best Local Similarity 12.2%; Pred. No. 0.00035;
Matches 31; Conservative 105; Mismatches 118; Indels 0; Gaps 0;

QY   195 AAAAGAGAAAAATAAAGAATGAGAGATAGAGAGATTTCTTATGAAAAAGAGAGAAC 254
```


Db 33 RARCRARURURARCRARURGRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 92
QY 255 TGTAGGTGAACAAATAAGAGATATGATATATTTATGAGAGGTGGTGAAGATTAT 314
Db 93 RNR 152
QY 315 TTTAGGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAAGATGA 374
Db 153 RNR 212
QY 375 ATTGTTGAAGATGAAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATG 434
Db 213 RNRNRRCRARGCRURGRURARARCRURCRURGRURGRURGRURGRURGRURGR 272
QY 435 AAAAAAGAAACAAA 448
Db 273 AAAAAAGAAACAAA 286

RESULT 9

US-07-922-723A-7

; Sequence 7, Application US/07922723A

; Patent No. 5369004

; GENERAL INFORMATION:

; APPLICANT: Drs. Mihael H. Polymeropoulos

; APPLICANT: and Carl R. Merrill

; TITLE OF INVENTION: FIVE HIGHLY INFORMATIVE

; TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe, Price, LeBlanc & Becker

; STREET: Suite 300, 99 Canal Center Plaza

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: DOS Text File

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/922,723A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: D.J. Mills

; REGISTRATION NUMBER: 34506

; REFERENCE/DOCKET NUMBER: 717081B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703 684 1111

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 291

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-07-922-723A-7

Query Match

Best Local Similarity 3.3%; Score 55.6; DB 1; Length 291;

Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAAGAAAGAGAACATGTAGGTGAACAAATAAGAGATATGATATATTTATG 296

Db 33 AGAAAGAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAA 92

QY 297 AGAGGTGTTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356

Db 93 AGGAAGGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152

QY 357 TGAATCTGAAGAAGATGAATTTGTTAAAGATGAAGAGAGAGAAAGAACTCCATGGCTAA 416
Db 153 AGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 212
QY 417 AGCTCTGTTAAAGACATGAAAAAGAAACAAAGAAAGAGAGAGAAAGAGAA 466
Db 213 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 262

RESULT 10

US-07-799-828C-7

; Sequence 7, Application US/07799828C

; Patent No. 5378602

; GENERAL INFORMATION:

; APPLICANT: Drs. Carl R. Merrill and

; APPLICANT: Mihael H. Polymeropoulos

; TITLE OF INVENTION: TWENTY SEVEN HIGHLY INFORMATIVE

; TITLE OF INVENTION: MICROSATELLITE REPEAT

; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS

; NUMBER OF SEQUENCES: 63

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe, Price, LeBlanc & Becker

; STREET: Suite 300, 99 Canal Center Plaza

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: DOS Text File

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/799,828C

; FILING DATE: 19911127

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: D.J. Mills

; REGISTRATION NUMBER: 34,506

; REFERENCE/DOCKET NUMBER: 717081A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703 684 1111

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 291

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-07-799-828C-7

Query Match 3.3%; Score 55.6; DB 1; Length 291;

Best Local Similarity 52.6%; Pred. No. 0.00065;

Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAAGAAAGAGAACATGTAGGTGAACAAATAAGAGATATGATATATTTATG 296

Db 33 AGAAAGAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAA 92

QY 297 AGAGGTGTTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356

Db 93 AGGAAGGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152

QY 357 TGAATCTGAAGAAGATGAATTTGTTAAAGATGAAGAGAGAGAAAGAACTCCATGGCTAA 416

Db 153 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 212

QY 417 AGTCTCGTTAAAGACATGAAAAAGAAACAAAGAAAGAGAGAGAAAGAGAA 466

Db 213 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 262

RESULT 11

```

US-08-074-275-7
; Sequence 7, Application US/08074275
; Patent No. 5468610
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,275
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/707,501
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-074-275-7

Query Match 3.3%; Score 55.6; DB 1; Length 291;
Best Local Similarity 52.6%; Pred. No. 0.00065;
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAGAAAGAGAGAACATGTTAGGTGAACAAAATAAAGAGATATGATATATTTATG 296
Db 33 AGAAGAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAAAGGA 92
QY 297 AGAGGTGGTGAAGATTATTTTAAAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152
QY 357 TGAATCTGAAGAGAGATGTTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416
Db 153 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 212
QY 417 AGTCTCGTAAAGAGATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 466
Db 213 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 262

RESULT 12
US-08-480-366-7
; Sequence 7, Application US/08480366
; Patent No. 5721100
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS

```

```

; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,366
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-480-366-7

Query Match 3.3%; Score 55.6; DB 1; Length 291;
Best Local Similarity 52.6%; Pred. No. 0.00065;
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAGAAAGAGAGAACATGTTAGGTGAACAAAATAAAGAGATATGATATATTTATG 296
Db 33 AGAAGAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAAAGGA 92
QY 297 AGAGGTGGTGAAGATTATTTTAAAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152
QY 357 TGAATCTGAAGAGAGATGTTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416
Db 153 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 212
QY 417 AGTCTCGTAAAGAGATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 466
Db 213 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 262

RESULT 13
US-07-952-277A-7
; Sequence 7, Application US/07952277A
; Patent No. 5861504
; GENERAL INFORMATION:
; APPLICANT: Drs. Michael H. Polymeropoulos
; APPLICANT: Drs. Carl R. Merrill
; TITLE OF INVENTION: ELEVEN HIGHLY INFORMATIVE
; TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match 3.3%; Score 55.6; DB 4; Length 44453;
Best Local Similarity 50.0%; Pred. No. 0.0026;
Matches 139; Conservative 0; Mismatches 139; Idels 0; Gaps

[illegible]

Search completed: November 28, 2002, 23:17:15
Job time : 816 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 20:43:06 ; Search time 268 Seconds

(without alignments)
14285.064 Million cell updates/sec

Title: US-09-733-685-3

Perfect score: 1700

Sequence: 1 ttgggtttttattgataaca.....tttaagacttaagacatat 1700

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.8	4.5	7657	22	Chemically pretrea
2	76.8	4.5	7657	24	Human immune syste
3	76.8	4.5	14006	24	Human immune syste
4	76	4.5	19124	18	Plasmodium var-7 g
5	76	4.5	19124	21	Plasmodium var-7 p
6	74.8	4.4	9539	22	Chemically pretrea
7	74.8	4.4	9539	24	DNA transcription
8	74.8	4.4	6644	20	Base sequence of t
9	74	4.4	7372	20	Base sequence of t

10	74	4.4	7797	20	AA333180	Cowpox virus bsr f
11	74	4.4	7996	20	AA333184	Base sequence of t
12	73.4	4.3	113515	24	ABL34174	Human immune syste
13	72.6	4.3	16033	24	ABL33404	Human immune syste
14	72	4.2	16167	24	ABL70254	Chemically treated
15	72	4.2	16167	24	ABL33083	Human immune syste
16	72	4.2	16167	24	ABL34529	Human metastasis a
17	71.6	4.2	17934	24	ABL33719	Human immune syste
18	71.2	4.2	10620	20	AA302996	Human IL-1ra BAC c
19	71.2	4.2	14690	20	AA322303	Human IL-1ra BAC c
20	70.6	4.2	34769	22	AA346775	Tumour suppressor
21	70	4.1	5750	22	AA346708	Tumour suppressor
22	70	4.1	5750	24	ABL34008	Human immune syste
23	69.6	4.1	556	24	ABQ36996	Oligonucleotide fo
24	69.6	4.1	556	24	ABQ36997	Oligonucleotide fo
25	68.4	4.0	723	24	ABQ25178	Oligonucleotide fo
26	68.4	4.0	723	24	ABQ25179	Oligonucleotide fo
27	68	4.0	292	24	ABL86605	Human ovarian canc
28	68	4.0	6145	24	ABL32972	Human immune syste
29	68	4.0	16633	24	ABN79984	Human chemically m
30	67.8	4.0	8079	24	ABL92313	Chemically treated
31	67.4	4.0	8404	22	AA346500	Tumour suppressor
32	67.4	4.0	8404	24	ABL33595	Human immune syste
33	67.2	4.0	556	23	ABV40063	Human prostate exp
34	67.2	4.0	556	23	ABV40163	Human prostate exp
35	67.2	4.0	556	23	ABV42105	Human prostate exp
36	67.2	4.0	556	23	ABV43601	Human prostate exp
37	67	3.9	6668	24	ABL36697	Ovary cancer relat
38	66.8	3.9	174424	22	ABL68122	Tumour suppressor
39	66.6	3.9	6485	24	AA346559	Human immune syste
40	66.6	3.9	6485	24	ABL33808	Human ovarian canc
41	66.4	3.9	299	24	ABL86618	Human prostate exp
42	66.4	3.9	563	23	ABV50724	Human immune syste
43	66.4	3.9	6338	24	ABL33126	Human immune syste
44	66.4	3.9	16200	24	AA336438	Mouse L66 genomic
45	66.2	3.9	307	24	ABL86438	Human ovarian canc

ALIGNMENTS

RESULT 1

AA345477/c

ID AA345477 standard; DNA; 7657 BP.

AC AA345477;

XX AA345477;

DT 18-DEC-2001 (first entry)

XX Chemically pretreated complementary DNA associated with cell cycle #91.

DE Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;

DE human immunodeficiency virus; neurodegenerative disorder; solid tumour;

DE graft-versus-host disease; glomerular disease; Lewy body disease; cancer;

DE arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;

DE immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;

DE PCR primer.

XX Homo sapiens.

OS WO200168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 30-JUN-2000; 2000DE-1019173.

PR 01-SEP-2000; 2000DE-1043529.

XX (EPIG-) EPIGENOMICS AG.

Dd	2568	TATCACATATAAACAATACTTAACAAAGAACCTTTTCACTTTAAAAATTCTCAATTTCTC	2509
Qy	145	ATCAAATCTAATCTCTTATATCATACAACGACATTTGACGGAAAATTCAGGTAAAAAGAGAAA	204
Dd	2508	ATCTACTATAAACAAATAA	2449
Qy	205	ATAAGAATGAGAGATAGAGAGATTTCTATGGAAAAGAAAGAGAGACATGTAGGTGA	264
Dd	2448	AAA	2389
Qy	265	CAAAATAAGAGATATCATGATATATTTTTATGAGAGTGCGTGAAGATTATTTTAGGAGAG	324
Dd	2388	AIAIAIACCATAA	2329
Qy	325	GGAGAGAAATAGAAAAGAAAATGACATGCTGAAGAAAGATGAATGTGTGTAA	384
Dd	2328	ATAA	2269
Qy	385	AGATGAAGAGAGAAAGAGACTCCATGGCTAAAGTCTCGTAAAGAGATGAAGAAAGAAC	444
Dd	2268	AAA	2209
Qy	445	AAAAAGAGAGAGAAAGAGAAAGGCTAAAAATAGACTAACTATTGCCAAA	496
Dd	2208	AAA	2157
RESULT 4			
AAT72882			
ID	AAT72882 standard; cDNA; 19124 BP.		
XX AC	AAT72882;		
XX DT	12-SEP-1997 (first entry)		
XX DE	Plasmodium var-7 gene.		
KW KW	DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy; Duffy binding like gene; Duffy antigen binding protein; erythrocyte; KWB; merozoite; malaria; var-1; var-2; var-3; var-7; immune response; Plasmodium; ss.		
OS OS	Plasmodium vivax. Plasmodium falciparum.		
XX FH	Key Location/Qualifiers		
FT exon	7317..15139 /*tag= a		
FT intron	/number= 1 15140..16205 /*tag= b		
FT FT	/number= 1 16206..17552 /*tag= C		
FT FT	/number= 2 /note= "no stop codon given"		
XX WO	9640766-A2.		
XX PN			
XX PD	19-DEC-1996.		
XX PF	07-JUN-1996; 96WO-US09508.		
XX PR	07-JUN-1995; 95US-0487826.		
XX PA	{USSH } US DEPT HEALTH & HUMAN SERVICES., Chitnis C, Miller LH, Peterson DS, Sim KL, Su X, Pellegrini TE;		
PI PI			
XX WPI	1997-052231/05.		
DR P-	PSDB; AAM22475.		
XX XX			

PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
XX binding proteins
XX Claim 4; Page 56-61; 96pp; English.
XX This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP), conserved regions (see AAT72889 and AAT72888 respectively). The
CC var family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and
CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the
CC invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
CC family of genes having homology with conserved regions of DABP and SABP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).
XX
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 18; Length 19124;
Best Local Similarity 50.0%; Pred. No. 1e-05;
Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
QY 52 ATTGGATGATGACAAATATTATATATACAGTGTAAATCAATCTTCTTAAAT 111
DB 15440 ATTTTAAATGAAAAAGAAAAATGAATATAAAAAAATTTATTAATAAATA 15499
QY 112 ACATGCTTTTAAATCAGACATTTGTTTAAATCAATCTTCTTAAAT 171
DB 15500 AAAAAAAGAAAAAGAGAAAAATTTTAAATAATTAATAAATAATA 15559
QY 172 GACATTGACGAAATTCAGGTAAAGAGAAAAATTAAGATGAGAGATTC 231
DB 15560 AATTTTGATAGATAAAAAATGAAAAAGATTTATCAAAAAAATTA 15616
QY 232 TATGGAAGAAAGAGAGACATGTAGGTGAACAAAAATTAAGAGATATGATAT 291
DB 15617 TATATAAAAAAATGATTTATTAATAAATAAAAAAGAAAAAATTA 15676
QY 292 TTATGAGAGGTGGTGAAGATTTATTTAGGAGAGGAGAGCAATAGAAAGAAATGA 351
DB 15677 AAAAAAATAATATATATATCAAAAAAAGAAAAAATTAATAATAA 15736
QY 352 CATGGTGAATCTGAAGAGATGATTTGTTTAAAGATGAAGAGAAAGAACTCCATG 411
DB 15737 AAATATATATCAATAAATAAAAAAATTAATAAATAATTAATAAATA 15796
QY 412 GCTAAAGTCTCGTAAAGAGATGAAAAAGAAACAAAGAGAGAGAAAGGC 471
DB 15797 AAATAAAAAAATTTATTAATAAATAAATAAATAAATAAATAAATTTAATAATAA 15856
QY 472 TAAATAGACTACTA 487
DB 15857 AAAAAAATAATAATA 15872

RESULT 5
AAZ98287
ID AAZ98287 standard; DNA; 19124 BP.
XX
AC AAZ98287;
XX
DT 13-JUN-2000 (first entry)
XX

DE Plasmodium var-7 polypeptide encoding DNA.
XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
XX protozoacide; var-7; ds.
XX Plasmodium sp.
OS
PN US5993827-A.
XX
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Sim KL, Chitnis C, Peterson DS, Su X, Wellem's TE, Miller LH;
XX WPI: 2000-194198/17.
DR P-PSDB; AAY77904.
XX
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
XX
XX Disclosure; Columns 91-108; 93pp; English.
XX
CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-7 polypeptide encoding
CC DNA.
XX
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 21; Length 19124;
Best Local Similarity 50.0%; Pred. No. 1e-05;
Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
QY 52 ATTGGATGATGACAAATATTATATATACAGTGTAAATCAATCTTCTTAAAT 111
DB 15440 ATTTTAAATGAAAAAGAAAAATGAATATAAAAAAATTTATTAATAAATA 15499
QY 112 ACATGCTTTTAAATCAGACATTTGTTTAAATCAATCTTCTTATATATCAAC 171
DB 15500 AAAAAAAGAAAAAGAGAAAAATTTTAAATAATTAATAAATAATA 15559
QY 172 GACATTGACGAAATTCAGGTAAAGAGAAAAATTAAGATGAGAGATTC 231
DB 15560 AATTTTGATAGATAAAAAATGAAAAAGATTTATCAAAAAAATTAATAATAA 15616
QY 232 TATGGAAGAAAGAGAGACATGTAGGTGAACAAAAATTAAGAGATATGATATAT 291
DB 15617 TATATAAAAAAATGATTTATTAATAAATAAAAAAGAAAAAATTAATAATAA 15676
QY 292 TTATGAGAGGTGGTGAAGATTTATTTAGGAGAGGAGAGCAATAGAAAGAAATGA 351
DB 15677 AAAAAAATAATATATATATCAAAAAAAGAAAAAATTAATAATAA 15736
QY 352 CATGGTGAATCTGAAGAGATGATTTGTTTAAAGATGAAGAGAAAGAACTCCATG 411
DB 15737 AAATATATATCAATAAATAAAAAAATTAATAAATAATTAATAAATA 15796
QY 412 GCTAAAGTCTCGTAAAGAGATGAAAAAGAAACAAAGAGAGAGAAAGGC 471

PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer
XX
PS Claim 1; SEQ ID No 54; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 other;

Query Match 4.4%; Score 74.8; DB 24; Length 9539;
Best Local Similarity 48.4%; Pred. No. 1.6e-05;
Matches 208; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 67 AAATATTAATATATCAGGTGTTAAATCAACATGTTTCTTTAAATACATGATTTTAAAA 126
DB 589 AAAAAAAAAATAAAAACAAAAAATAAAAAAACAACAAAAAACAACAAAAA 530
QY 127 TCAGACATTTGTTTAAATCAATCTATCTTTATATCAGACGATTCAGCGAAAA 186
DB 529 AAAAAAAAAATCAAAAAAATAAAAAAATAAAAAAACAACAAAAAACA 470
QY 187 TTCAGTAAAGAGAAAAATAAGAAATCAGAGATAGAGAGATTTCTATGGAAGAAAG 246
DB 469 TAAAAAATAAATAAAAAAATAAAAAAATAAAAAAACAACAAAAAATAAAAA 410
QY 247 AGAGAACATGTAGGTGAACAAATAAAGAGATATGATGATATATTTATGAGAGGTGGT 306
DB 409 AAAAAAACGAAAAAACAACAAAAAATAAAAAAATAAAAAAACAACAAAAAATAAAAA 350
QY 307 AGATATTTTGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAA 366
DB 349 AAAAAAATAAATAAAAAAATAAAAAAATAAAAAAACAACAAAAAATAAAAA 290
QY 367 GAAGATGATTTGTTTAAAGATGAAGAGAGAGAAAGAACTCCATGGCTTAAAGTCTCTGTA 426
DB 289 AAAAAACAACCAAAAAAATAAAAAAATAAAAAAATAAAAAAACAACAAAAAATAAAAA 230
QY 427 AGAAGATGAAAAAGAAACAAAGAGAGAGAGAGAAAGGCGTAAAAATAGACTACT 486
DB 229 ATAAACGAAAAATACAAAAAATAAAAAAATAAAAAAATAAAAAAACAACAAAAAATAAAAA 170
QY 487 ATTGCCAAAA 496
DB 169 AAAAAAATAA 160

RESULT 8
AAK33181
ID AAK33181 standard; DNA; 6644 BP.
XX

AC AAK33181;
XX 25-JUN-1999 (first entry)
DT
XX
DE Base sequence of the plasmid pRx-ires-bsr.
XX
XX Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
OS Synthetic.
OS Cowpox virus.
XX
XX WO9913073-A2.
PN
XX 18-MAR-1999.
PD
XX 07-SEP-1998; 98WO-JP04010.
PF
XX 08-SEP-1997; 97JP-0259235.
PR
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PA
XX Hamada H;
PI
XX WPI; 1999-243728/20.
DR
XX New apoptosis-resistant virus-sensitive cell
PT
XX
XX Example 1; Page 38-41; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC is used in an example from the present invention.
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;
Query Match 4.4%; Score 74; DB 20; Length 6644;
Best Local Similarity 48.4%; Pred. No. 2.1e-05;
Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
QY 107 AAATACATGCAATTTTAAATCAGACATTTGTTTAAATCAATCTATCTTTATATC 166
DB 4018 AA 4077
QY 167 ACAACGACATTCAGCGAAATTCAGGTAAAGAGAGAAATAAAGATGAGATAGAGAG 226
DB 4078 AA 4137
QY 227 ATTTCTATGAAAAAGAAAGAGAGACATCTAGGTGAACAAATAAAGAGATATGATGAT 286
DB 4138 AA 4197
QY 287 ATATTTTATGAGGTGTCAGATTTATTTAGGAGGGGAGAGAGAAATAGAAAAAGAA 346

Db 4198 AA 4257
 QY 347 AATGACATGGTCAATCTGAAGAGTGAATTTGTTAAAGATGAAGAGAGAACT 406
 Db 4258 AA 4317
 QY 407 CCATGGCTAAAGTCTGCTAAAGAGATGAAAGAGAAACAAAGAGAGAGAGAGAGA 466
 Db 4318 AA 4377
 QY 467 AAGGCTAAATAGACTAATCTGTCACAAATTTCTGTAGCGCAGCAATATCTTTGGTCC 526
 Db 4378 AA 4437
 QY 527 AAGGTT 532
 Db 4438 AAGCTT 4443
 RESULT 9
 AAX33182
 ID AAX33182 standard; DNA; 7372 BP.
 XX
 AC AAX33182;
 DT 25-JUN-1999 (first entry)
 XX
 DE Base sequence of the plasmid prX-Bcl-xI-bsr.
 KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
 KW crmA; bcl-2; bcl-xI; FLIP; survivin; IAP; ILP; adenovirus; cancer;
 KW autoimmune disease; graft rejection reaction; inflammation;
 KW inflammatory disease; ss.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO9913073-A2.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98WO-JP04010.
 XX
 PR 08-SEP-1997; 97JP-0259235.
 XX
 PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
 XX
 PI Hamada H;
 XX
 WPI; 1999-243728/20.
 XX
 PT New apoptosis-resistant virus-sensitive cell
 XX
 PS Example 2; Page 41-45; 51pp; English.
 XX
 CC The present invention describes an apoptosis-resistant virus-sensitive
 CC cell line into which an apoptosis resistance gene has been introduced.
 CC The recombinant viruses generated are capable of expressing apoptosis-
 CC associated genes. These can then be used in a variety of diseases for
 CC which the induction of apoptosis by gene transfer, or where the
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
 CC are useful as vectors for gene therapy which can be applied to cancer
 CC therapy for destroying cancer cells selectively, the treatment of
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction
 CC therapy for inflammatory cells in inflammatory reaction, and apoptosis induction
 CC encountered the problem where if an adenovirus vector capable of
 CC expressing an apoptosis-associated gene is introduced into animal cells,
 CC the cells producing the virus will be destroyed because the period of
 CC time required to induce cell death by apoptosis is shorter than that
 CC required to replicate and produce the virus, resulting in failure to
 CC obtain a recombinant virus having the integrated apoptosis-associated
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an

CC apoptosis resistant gene introduced) is established and overcomes the
 CC problem. The present sequence represents the base sequence of the
 CC plasmid prX-Bcl-xI-bsr, which contains the human Bcl-xI gene, and
 CC is used in an example from the present invention.
 XX
 SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 Other;
 Query Match 4.4%; Score 74; DB 20; Length 7372;
 Best Local Similarity 48.4%; Pred. No. 2.2e-05;
 Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
 QY 107 AATAATCATGCAATTTAAATCAGACATTTGTTTTTAAATCAATCTCTTATATC 166
 Db 4746 AA 4805
 QY 167 ACAACGACATTCAGGAAATTCAGGTAAAGAGAGAAATTAAGATGAGAGATAGAGAG 226
 Db 4806 AA 4865
 QY 227 ATTTCTATGCAAAAGAGAGAGACATGTAGGTCAACAAATATAAGAGATATGATGAT 286
 Db 4866 AA 4925
 QY 287 ATATTTTATGAGAGGTGTCAGATTTATTTTAGGAGGGAGAGAGAGAGAGAGAGAG 346
 Db 4926 AA 4985
 QY 347 AATGACATGGTGAATCTGAAGAGATGAATTTGTTAAAGATGAAGAGAGAGAGAGAG 406
 Db 4986 AA 5045
 QY 407 CCATGCTAAAGTCTGCTAAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 466
 Db 5046 AA 5105
 QY 467 AAGGCTAAATAGACTAATCTGTCACAAATTTCTGTCAGCGCAATATCTATTTGGTCC 526
 Db 5106 AA 5165
 QY 527 AAGGTT 532
 Db 5166 AAGCTT 5171
 RESULT 10
 AAX33180
 ID AAX33180 standard; DNA; 7797 BP.
 XX
 AC AAX33180;
 DT 25-JUN-1999 (first entry)
 XX
 DE Cowpox virus bsr full length gene sequence.
 XX
 KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
 KW crmA; bcl-2; bcl-xI; FLIP; survivin; IAP; ILP; adenovirus; cancer;
 KW autoimmune disease; graft rejection reaction; inflammation;
 KW inflammatory disease; ss.
 OS Cowpox virus.
 XX
 PN WO9913073-A2.
 XX
 PD 18-MAR-1999.
 XX
 PF 07-SEP-1998; 98WO-JP04010.
 PR 08-SEP-1997; 97JP-0259235.
 XX
 PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
 XX
 PI Hamada H;
 XX

DR WPI; 1999-243728/20.
 XX New apoptosis-resistant virus-sensitive cell
 PT Example 1; Page 34-38; 51pp; English.
 XX
 PS The present invention describes an apoptosis-resistant virus-sensitive
 CC cell line into which an apoptosis resistance gene has been introduced.
 CC The recombinant viruses generated are capable of expressing apoptosis-
 CC associated genes. These can then be used in a variety of diseases for
 CC which the induction of apoptosis by gene transfer, or where the
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
 CC are useful as vectors for gene therapy which can be applied to cancer
 CC therapy for destroying cancer cells selectively, the treatment of
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
 CC encountered the problem where if an adenovirus vector capable of
 CC expressing an apoptosis-associated gene is introduced into animal cells,
 CC the cells producing the virus will be destroyed because the period of
 CC time required to induce cell death by apoptosis is shorter than that
 CC required to replicate and produce the virus, resulting in failure to
 CC obtain a recombinant virus having the integrated apoptosis-associated
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an
 CC apoptosis resistant gene introduced) is established and overcomes the
 CC problem. The present sequence represents the cowpox virus bsr gene which
 CC is used in an example from the present invention.
 XX
 SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;

Query Match 4.4%; Score 74; DB 20; Length 7797;
 Best Local Similarity 48.4%; Pred. No. 2.2e-05;
 Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
 QY 107 AAAATACATGCAATTTAAATCAGACATTTGTTTAAATCAATCTTATATC 166
 Db 5171 AA 5230
 QY 167 ACAACGACATTCAGGAAATTCAGTAAAGACAGAAATTAAGATCAGAGATAGAGAG 226
 Db 5231 AA 5290
 QY 227 ATTTCTATGCAAAAGAGAGAGACATGTAGGTGAACAAATTAAGAGATATGATGAT 286
 Db 5291 AA 5350
 QY 287 ATATTTTATGAGAGGTGTCAGATTTATTTAGGAGAGGAGAGAGAAATGAGAAAGAA 346
 Db 5351 AA 5410
 QY 347 AATGACATGTGTAATCTGAAGAGATGTAATTTGTTTAAAGATGAAGAGAGAACT 406
 Db 5411 AA 5470
 QY 407 CCATGGCTAAGTCTCGTAAAGAGATGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAG 466
 Db 5471 AA 5530
 QY 467 AAGGCTAAATACACTATGTCACAAATTTCTGACCCGACAAATACTATTGTCGCC 526
 Db 5531 AA 5590
 QY 527 AAGGTT 532
 Db 5591 AAGCTT 5596

RESULT 11
 AAX33184
 ID AAX33184 standard; DNA; 7996 BP.
 XX
 AC AAX33184;
 XX
 DT 25-JUN-1999 (first entry)

XX DE Base sequence of the plasmid pRx-Bcl 2-i-hcd 25.
 XX
 KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
 KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
 KW autoimmune disease; graft rejection reaction; inflammation;
 KW inflammatory disease; ss.
 XX Synthetic.
 OS Homo sapiens.
 XX WO9913073-A2.
 XX 18-MAR-1999.
 XX 07-SEP-1998; 98WO-JP04010.
 XX 08-SEP-1997; 97JP-0259235.
 XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
 XX Hamada H;
 XX WPI; 1999-243728/20.
 XX New apoptosis-resistant virus-sensitive cell
 XX
 XX Example 3; Page 46-49; 51pp; English.
 XX The present invention describes an apoptosis-resistant virus-sensitive
 CC cell line into which an apoptosis resistance gene has been introduced.
 CC The recombinant viruses generated are capable of expressing apoptosis-
 CC associated genes. These can then be used in a variety of diseases for
 CC which the induction of apoptosis by gene transfer, or where the
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
 CC are useful as vectors for gene therapy which can be applied to cancer
 CC therapy for destroying cancer cells selectively, the treatment of
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
 CC encountered the problem where if an adenovirus vector capable of
 CC expressing an apoptosis-associated gene is introduced into animal cells,
 CC the cells producing the virus will be destroyed because the period of
 CC time required to induce cell death by apoptosis is shorter than that
 CC required to replicate and produce the virus, resulting in failure to
 CC obtain a recombinant virus having the integrated apoptosis-associated
 CC gene. In this invention an apoptosis-resistant 293 cell line (having an
 CC apoptosis resistant gene introduced) is established and overcomes the
 CC problem. The present sequence represents the base sequence of the
 CC plasmid pRx-Bcl 2-i-hcd 25, which contains the human Bcl-2 gene, and
 CC is used in an example from the present invention.
 XX
 SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other;

Query Match 4.4%; Score 74; DB 20; Length 7996;
 Best Local Similarity 48.4%; Pred. No. 2.2e-05;
 Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
 QY 107 AAAATACATGCAATTTAAATCAGACATTTGTTTAAATCAATCTTATATC 166
 Db 5370 AA 5429
 QY 167 ACAACGACATTCAGCGGAAATTCAGGTAAAAAGAGAAATTAAGATGAGAGATAGAGAG 226
 Db 5430 AA 5489
 QY 227 ATTTCTATGCAAAAGAGAGAGACATGTAGGTGACAAATTAAGAGATATGATGAT 286
 Db 5490 AA 5549
 QY 287 ATATTTTATGAGAGGTGTCGTAAGATTTATTTAGGAGAGGAGAGAGAAATAGAAAGAA 346
 Db 5550 AA 5609

XX Human immune system associated gene SEQ ID NO: 1056.
 DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 1056; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 16167 BP; 3894 A; 285 C; 4136 G; 7838 T; 14 other;
 Query Match 4.2%; Score 72; DB 24; Length 16167;
 Best Local Similarity 46.7%; Pred. No. 6.3e-05;
 Matches 207; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
 QY 25 AAATATTATTATTCATGAGTTTATTTGGATAGCATGACAAATATTAAATATATATCAGT 84
 DB 5977 AATAATATAAAATCTTAATTAATTTATAAAATTTAACTTATTAATTAATTTTAAAAAT 5918
 QY 85 GTTAATAACATGTTTCTCTTAAATACATGCAATTTTAAATCAGACATTTGTTTAA 144
 DB 5917 TAATTAATTTCAATTTATTAATCTTAACTTTAAATATACCTCATTTTATTTCTACT 5858
 QY 145 ATCAAAATCTAATCTTTATATCATCAACGACATTCAGGAAATTCAGGTAATAAGAGAAA 204
 DB 5857 TTTCTCACACCAAAACGNNNAACAACGAAACAAACAAACAAACAAACAAACAAACAA 5798
 QY 205 ATAAGAATGAGAGATGAGAGATTTCTATGGAAAGAAAGAGAGAAACATGTAGGTGAA 264
 DB 5797 AACGAAACAAACAAACAAACAA 5738
 QY 265 CAAATAAGAGATATCATATATTTTATGAGAGGTGCTGAAGATTATTTTAGGAGAG 324
 DB 5737 ACGANAAAAACGAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 5678
 QY 325 GGAGAGAGAAATAGAAAAAGAAATGACATGCTGAAGAGATGAATTTGCTTAA 384
 DB 5677 AACGCGCAACAAAAA 5618

QY 385 AGATGAAGAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATCAAAAAAGAAAC 444
 DB 5617 AAA 5558
 QY 445 AAAAGAGAGAGAGAGAGAAAGAGAA 467
 DB 5557 AAAANAAACACAAAAA 5535

Search completed: November 28, 2002, 23:12:13
 Job time : 805 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 21:46:16 ; Search time 3105 Seconds
(without alignments)
15933.903 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....tttaacgacttaacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vt:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rdt:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1700	100.0	83650	8	AB023041	Arabidops
C 2	576.2	33.9	92620	8	AB026636	Arabidops
C 3	540.2	31.8	83646	8	AB005248	Arabidops
C 4	534.6	31.4	94487	8	AC012394	Arabidops
C 5	534.6	31.4	100806	8	AC015450	Arabidops
C 6	499.8	29.4	104386	8	ATP32A17	Arabidops
C 7	499.8	29.4	179771	8	ATCHRIV25	Arabidops
C 8	497	29.2	95519	8	AF071527	Arabidops
C 9	497	29.2	116448	8	AC005142	Arabidops
C 10	497	29.2	159629	8	ATCHRIV9	Arabidops
C 11	369	21.7	369	8	ATH250185	Arabidops
C 12	330.8	19.5	95190	8	AC007203	Arabidops
C 13	103.2	6.1	7218	6	166494	Sequence 14
C 14	90.2	5.3	171763	2	AC115878	Mus muscu
C 15	89.2	5.2	280982	2	AC125184	Mus muscu
C 16	87.4	5.1	161362	2	AC101939	Mus muscu
C 17	87.4	5.1	217779	2	AL844515	Mus muscu
C 18	86.8	5.1	202872	2	AC016160	Homo sapi
C 19	86.6	5.1	166299	2	AC127244	Mus muscu
C 20	85.6	5.0	44735	9	AC092315	Homo sapi
C 21	85.6	5.0	48623	2	AC130469	Homo sapi
C 22	85	5.0	178038	10	AL672308	Mouse DNA
C 23	84.6	5.0	422	5	PD0004	P.domesticu
C 24	84.6	5.0	61953	2	AC101966	Mus muscu
C 25	84.6	5.0	177648	10	AL732328	Mouse DNA
C 26	84.6	5.0	228944	2	AL844530	Mus muscu
C 27	84.4	5.0	222556	2	AC126558	Mus muscu
C 28	84	4.9	417	5	AF044763	Cecropia
C 29	84	4.9	202756	2	AC119804	Mus muscu
C 30	84	4.9	218898	2	AC114903	Mus muscu
C 31	83.6	4.9	162803	2	AC117818	Mus muscu
C 32	83.6	4.9	206538	2	AC102446	Mus muscu
C 33	83.4	4.9	239851	2	AC101712	Mus muscu
C 34	83.4	4.9	247340	2	AC121591	Mus muscu
C 35	83.2	4.9	64789	2	AC083839	Homo sapi
C 36	82.8	4.9	167364	10	AC121590	Mus muscu
C 37	82.8	4.9	220275	10	AL732404	Mouse DNA
C 38	82.6	4.9	48699	2	AC115677	Dictyoste
C 39	82.4	4.8	178413	2	AL805954	Mus muscu
C 40	82.4	4.8	203295	10	AL663049	Mouse DNA
C 41	82.2	4.8	204829	2	AC113318	Mus muscu
C 42	82	4.8	163132	2	AC097180	Rattus no
C 43	81.6	4.8	93661	9	AC107463	Homo sapi
C 44	81.6	4.8	106615	10	AL645809	Mouse DNA
C 45	81.6	4.8	120635	2	AC127463	Ornithorh

ALIGNMENTS

RESULT 1
AB023041/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

AB023041 83650 bp DNA linear PLN 27-DEC-2000
Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: MPE11.
AB023041 BA000014
AB023041.1 GI:4220640

Arabidopsis thaliana (strain:Columbia) DNA, clone:Lib:Mitsui Pl
clone:MPE11.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.

REFERENCE

AUTHORS

TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. I.	CDS	SWRFEFLPAE"
JOURNAL	Sequence features of the regions of 4,504,864 bp covered by sixty		complement(join(10090..10408,10649..10804,10916..10962))
MEDLINE	PI and TAC clones		/note="gene_id:MPE11.3
REFERENCE	DNA Res. 7 (2), 131-135 (2000)		similar to unknown protein
AUTHORS	20277480		sp P42744"
TITLE	2 (bases 1 to 83650)		/codon_start=1
JOURNAL	Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.		/evidence=not_experimental
	Direct Submission		/protein_id="BAB01049.1"
	Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research		/db_xref="GI:9279591"
	Institute, Department of Plant Gene Research; 1532-3, Yana,		/translation="MMSPKAKYDQLMYTTOGTLEBASICLLNCGPISSNALKNVVLG
	Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,		GVGSITIVGSKVLGIDWKFQHRHAIEQKESISEGFRDENTVFORRHQSVFQRL
	Tel:81-438-52-3935, Fax:81-438-52-3934)		BQNRAGQTVPRMDIARRIWASICRMWMSLADRTTYTCKEARAITDPPFGRVLRITSL
COMMENT	Address for correspondence: kaos@kazusa.or.jp		VDRLDFRLCDI"
	For the latest information on annotation of this clone, please see		complement(11450..11818)
	http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=Mpe11		/note="gene_id:MPE11.4
	Genes with similarity to proteins in the databases are described in		unknown protein"
	'product' or 'note' qualifiers. Genes that have no significant		/codon_start=1
	protein similarity are described as 'unknown protein'.		/evidence=not_experimental
	The software programs used to predict genes include: Grail		/protein_id="BAB01050.1"
	(Informatics Group, Oak Ridge National Laboratory,		/db_xref="GI:9279592"
	http://combio.ornl.gov/Grail-1.3/).		/translation="MNSLKKERVEEDNCKSDGNRCKPSTEVVTVTVEEVDPPKI
	GENSCAN (Chris Burge, MIT, http://CCF-081.mit.edu/GENSCAN.html),		LRRVHATRVAKVNGYVAGELPSPKRRKRSQNLGRNSLDGNGVDGDFEINRVGL
	NetGene2 (S.M. Hebsgaard, et al., CSS, Technical University of		QGLGIDLNCRPEPDSVLSL"
	Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and		16981..17979
	SplicePredictor (Volker Brendel, Stanford University,		/note="gene_id:MPE11.5"
	http://grenlini.2001.iastate.edu/cgi-bin/sp.cgi).		/codon_start=1
	Genes encoding tRNAs are predicted by tRNAscan-SE		/evidence=not_experimental
	(Sean Eddy, Washington University School of Medicine, St. Louis,		/product="AP2 domain transcription factor-like protein"
	http://genome.wustl.edu/eddy/tRNAscan-SE/).		/protein_id="BAB01051.1"
	This sequence may not be the entire insert of this clone. It may be		/db_xref="GI:9279593"
	shorter because we remove overlaps between neighboring submissions.		/translation="MAERKRSSIQTNKPNKPMKKPPQNLNHLPLGSLDLMTKRLR
	The 5' clone is K9122 and the 3' clone is MLL14.		FVNDVATDIYSSSEERQRKRYVCEIDLFFQAQTAQAESESYCOEENNNGVSK
FEATURES	Location/Qualifiers		TK1SACSKVLRSAKSPVGVRSSTTSKPVGRQKRWKAAEIRHPIKTVTLGT
source	1..83650		ETLEQAADAVATKKLEDAALAAATSAASSVLSNCSGMSIASSGSSIDLKLVDSFLD
	/organism="Arabidopsis thaliana"		QAGESKAKSPDFADLOIPMGCFIDDSFIPNACELDFLLTEENNLMDDYCGID
	/strain="Columbia"		DDLIIGLECDGSPDPYDFSDVEIDLGLLGTIDKAVFVDHATTTPTPLNATCP"
	/db_xref="taxon:3702"		join(21893..22063,22330..22386,22578..22783,22894..23227,
	/chromosome="3"		23311..23382,23473..23613)
	/clone="MPE11"		/note="gb AAAF23821.1
	/clone_lib="Mitsui Pl"		gene_id:MPE11.6"
	complement(542..764)		/codon_start=1
exon	/note="CDS is reported in Acc# AP000599		/evidence=not_experimental
	contains similarity to CHP-rich zinc finger protein		/product="homocysteine S-methyltransferase AthMT-1"
	gene_id:K9122.5"		/protein_id="BAB01052.1"
	/number=1		/db_xref="GI:9279594"
	/evidence=not_experimental		/translation="MVLEKKSALLEDLIKKCGCAVVDGFGFNTQLEIHAANNDPLWS
	join(1705..2463,2548..2739,2862..2949,3037..3107,		AVSLIKPELIKRVHSEYLEAGDIVVTSSYQATIPFLSRGLSIEESLQKSVEL
	3190..3273,3410..3514)		AVEARDREKWSKVSCHYNRAVVAASIGSYGAYLADGSEYSGHYGENVSLDKMDF
	/note="gb AAD55139.1		HRRLOVLVEAGPDLLAFETIPNKLQAQACVELLEEKVQIPAWICFTSVGDEKAPSG
	gene_id:MPE11.1"		ESFECLPLNKSNNIYAVGINCAPQFENLIRKFAKLTKRAIVVYPSNGEWDGKA
	/codon_start=1		KOWLPSCGDDREFEATKWDIGAKLIGCCRTTPTSTINATSRDLKRR"
	/evidence=not_experimental		complement(2777..27552)
	/product="dihydrolipoamide S-acetyltransferase"		/note="gene_id:MPE11.7
	/protein_id="BAB01047.1"		unknown protein"
	/db_xref="GI:9279589"		/codon_start=1
	/translation="MTVRSKIREIFMPALSTMTGKIVSWIKTEGEKLAKGVSVVV		/evidence=not_experimental
	ESDKADMVETFDYGLAAIVVGEGETAPVGAALGLIAETEAEIEAKSKAASKSSSS		/protein_id="BAB01053.1"
	VAEAVPVPPTSPAPAIQAQPAVTSVSDGPKRTVATPYAKKLAKQHKVQIESVAG		/db_xref="GI:9279595"
	TGPFGRITASDVETAAAGTAPSKSIAPPPPPPVATKATTNLPPLPDSIVPFTA		/translation="MTHAREWRSLLTTLMLVLLSYMLGLPCVSRVCAIRIEPETP
	MQSAVKMIESLSVPTFVGVYVNTDALLALEKYPKGYWTALLAKAAGMALAQH		ASGKQEDLMKKYFGAGKFPVDSFVGKGISSEKRVISCPDPLN"
	PVYNASCKDGKSFSTNINAVAVANGLLTIPVQADKLDLILLSQKRELVGKA		30765..31883
	RSKQLOPHEYNSGFTSNLGMFGVDRFDAILPPQGAIMAVGASKPTVVVADKDGFFS		/note="gb AAFI6548.1
	VKNTMLVNTADHRIVYGAADLAFLQTFAKIENPDLSLT"		gene_id:MPE11.8
	complement(4594..5106)		similar to unknown protein"
	/note="gene_id:MPE11.2		/codon_start=1
	unknown protein"		/evidence=not_experimental
	/codon_start=1		/protein_id="BAB01054.1"
	/evidence=not_experimental		/db_xref="GI:9279596"
	/translation="MEDLLEERLRTDVSNGKRVKRGDLSDVVKRLRDLDFDSSL		/translation="MPKERRSVSLDKYKRSPLCCEASLAKPSQKQKWEERACP
	DPVSDQLDSVMKSFENELSTTAAALSGSETQPOLGYFEASDDELGLPPLTPQTL		NCMPHPINGILLICSSYNGCRPYMCDTSHRNSCNCFDQFRKASKERPSLSLREERES
	PPSCBETVTELVRASDSSSEVGLCGFEDHVTETFGPCDLDGDLFEYFEDGCLDSDLF		VEPTMEDVSDSTAVNLGGAASEITVVDSDGERGEVEVEVEVEVEVEVEVEEIV
			TTEODQKNKPOKLTCPLCRGHIKEWVVKAAFCMNSKHRSCSCCTCFSGYSOLR
			KHARLLHPGVVRPSEADPERQSWRRLERQSDLDGLLSTLQSSFGGDEISDNDGFLFAD


```

Qy 1501 GSCCTTCCCTGGCCGCTTTAGTAATATTCAGAGTAGGATAGGCTATGGCGCTGATAGGGAGT 1560
|||||
Db 11180 GCGCTTCCCTGGCCGCTTTAGTAATATTCAGAGTAGGATAGGCTGATAGGGAGT 11121
|||||
Qy 1561 CCAAAAAGAGCGGCGATTCGCGGTTGGGTGGCTTTGGAACTTTGGATTGCGATTAGT 1620
|||||
Db 11120 CCAAAAAGAGCGGCGATTCGCGGTTGGGTGGCTTTGGAACTTTGGATTGCGATTAGT 11061
|||||
Qy 1621 CATGTTTATCTATTAAATGCTCGCGACTTCGTGACGACGCGCTGTGTTCTTCTTCCCTCTG 1680
|||||
Db 11060 CATGTTTATCTATTAAATGCTCGCGACTTCGTGACGACGCGCTGTGTTCTTCTTCCCTCTG 11001
|||||
Qy 1681 TTTACGACTTACGACATAT 1700
|||||
Db 11000 TTTACGACTTACGACATAT 10981
|||||

RESULT 2
AB026636 92620 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K14A17.
ACCESSION AB026636 BA000014
VERSION AB026636.1 GI:4757392
KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
SOURCE clone:K14A17.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
p1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
2 (bases 1 to 92620)
Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
Direct Submission
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K14A17
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlin1.vol.tastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MH15 and the 3' clone is MCE21.
Location/Qualifiers
1..92620
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K14A17"
/clone_lib="Mitsui TAC"
complement(201..2180)
CDS

```

```

/note="gene_id:K14A17.1"
/codon_start=1
/evidence=not_experimental
/product="salt-inducible protein-like"
/protein_id="BAA94973.1"
/db_xref="GI:7670019"
/translation="MRGPASSASRIATAAFAKSLNAS"SYNPKLSKTLNSSGKPTNP
LNQYISOVIERKDWFLINQEFTHRIGLNTREVISV QNQNPLHSIRFYLWVSNF
APYKADQSLKSLGNALFRKGPLLSMELLKLNDSRISDBLMKVGLGSGWRGLG
DKVNDVPAQJSELGMPKSTRLYNAYIDALVKSNSLDLVLRFHOMSDCKPDRFTY
NILLHGCKGQVDEARLVKQMEQGNRPNI IYTLIDGFLACRVDEALKOLEMM
RVRLNPLEATIRTFVHGIFRCLPCKAFELVGMER SNLQVSGVDVAVLYCUNNS
MAKETQGFRLIGERYTPDSFNAAMSCLLKGEDLVTCRIFDSVRGKPGFNG
YLVLVALLNAQRFSEGRYDKOMGVGLLSVSYNAVINCLCKARRIENAAFLFE
MODRGISPLNVTENTFLSGYSVRGDKVKGHVLEKLLV GFKPDVTFSLINCLRA
KEIKDAFDCFKEMLEWLEPNEITYNLIIRSCSTGDI RSVKLFKAKENGSLSPDLV
AYNATIQSCFKMRVKKABELLKTMLRIGLCKGNTFY LKALSSEGRESEAREMS
SIERHGCVDSYTRKLVVEELDRLKSGLSRTVSA"
Join(3906..4124,4238..4462,4536..4649)
/note="embI: CAB82745.1"
gene_id:K14A17.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94974.1"
/db_xref="GI:7670020"
/translation="MAKQDPTSSNTMLPLVGEI EVRFQELDTTGTVDKATGNKVF
QKTPFLGLHNKRILVDPNDSPITVMKMKVTSFHRDCA YRGSOLDKDIPTVRKSSIV
OLKTRVEFLKHNOTRESSCPTIKGRMKRAITTYVALSTKITAQVYEGHERIVATI
YPNDYAIRIVLILFIEDLINMGTCI"
complement(join(5031..5929,6011..5601,6984..7211))
/note="gene_id:K14A17.3"
/codon_start=1
/evidence=not_experimental
/product="AMP-binding protein"
/protein_id="BAA94975.1"
/db_xref="GI:7670021"
/translation="MAATPKWRDIDDLPKIPANITALTPLWFLDRAAVVHPTKRSVING
SREYTRQYDRCRLASALADISGPGSIVFVALTANIPAMYEAFHGVPMCGAVIN
CVNRLNAPTVAFLHSQSSVIMVDQEFITLHETLMEERAGSFPKRLILVIGD
HTCAPESLNRAISKGAIEYEDFLATGDPNIPWCPHAFIPLQSTALGYSQTTSKPGVV
LHKGAVIMALSNSPLIMGMOGAVYLATLMEHCLMVC PUSLAVISGTSICLRGVV
KEYVSNIAKYKTHFCAPAVYLVNAILVNPKEI ILIIPVHVWVTGAAPPPSLVFSM
NOKGRVATYGLSETYGPSTVCAWKPEWDSI DEICANQAGVTHPTGHEQDVID
TGTGKVPADGATAGEIVFRGNVMVGYLKNPI AN-ET AGGWFFSHDIAYKHDPNDI
ETIKRSKDVIIISGENISSEVENVYVHPPAVLEASVAVRDPWQSRSCAFVFLKSD
YEKHONKLAQDINKFCREKLPAYVWPKSVVFLPKRIATGKIQKILHATKAKEMGPV
PRSLI"
complement(join(8463..8892,8977..9336,9230..9641))
/note="gene_id:K14A17.4"
/codon_start=1
/evidence=not_experimental
/product="basic chitinase"
/protein_id="BAA94976.1"
/db_xref="GI:7670022"
/translation="MVSRLPFLSLLLTVALVVIQGLVNAEDSEPSSTRKPLVKIV
KKGKLCRQWCKGMSYCCNHTISDFETQFENLFSKRNSVAHAFVGFWDYSFIF
AAAEYQPLGFGTAGKLGKMEVAFAFGLHGVSKTSCGYGATGPGFLAWGLKEMSP
DQLYCDDYYKLTYPCTPGVSYHGGALPVYWNYNVGTGEALKVDLLSHPEYLENNAT
LAFQALWVWMTPPKHLPLPSADHVEVGKWKPTKNDTAARTPGFGATINVLYGQICN
SGFDNMNNIVSHLYILDLIGVREAGPHEKLSCADQEPFSSSSAPSSSSSS"
Join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/note="gene_id:K14A17.5"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translation="MDGGLGLIGLIGSETHGFHTLQDLVDVTLBEAKSRWLRPNEIHA
ILCGRIILDFDRKMLNFRKDGHNKKKDKGRTVKEAHEHLKVGNIEERHLYVAGEDN
TTFVRRCYWLIDKARENILVHVHRTQEAHTSGDISSPISVSISVFPNVAADID

```

TVVRNHDLSLHDINTLDWDELLVPTDLNNOSAPTVDNLNLSYFTEPLQNAANGTAEGHNA
TVADGSLDALLNDGQPSRESFGRWMSFSESNGSLDPSFEPWVMPRODPLAPQAVF
HSHSNFQEVNTVDSFWAYSESSEKTLVLTGFLHDSYOHLSNLYCVGDFCVPA
EYLAQGVYRCIIPHSFGMVNLYLSADGHKPIQCFRFEHRAVPLDKTVPEDNQDSK
WEEFQVRLSHLFTSSNKNLSSKIPSHNLNDAKTLKASKTNHLLNSWALVKSIO
GNKVSFOADHLFLFSLKNRLKWLMEKLEGRNLDYDLSKGLGVHLCASIGYIWS
VQLFSLGSLNFRDQGWALHWAAYGREKNVAALISAGARPNIATVTDKNDLGGC
MAADLAONGYDGLAAYLAEKVLAQFQDMKIAGNITGDLACKAEMLNQGTLPDEQ
SLKDALAYTAEEAARIQAFREKALKARSVIOFANKKEEKSIIAAMKIQAF
RYKDRKKIAATRIQCFQTKWIRREYLMNRQALRIQAFFRGLQARQYKRLIWSV
FLKLEKAVRWQRKFRGLOVAEEDSPEAQEDFYKTSQRAEERLERSVVRVQAM
FRSKKAQDYYRMRKLTHEEAQVNHFLNLFSFGKKNRR"
complement(join(18170, .18172,18407, .18490,18736, .18807,
18977, .19090,19168, .19281,19387, .19487,19590, .19668,
19755, .19805,19906, .19989,20365, .20431,20539, .20669,
20916, .21040,21126, .21231,21349, .21483,21624, .21770,
21860, .22171))
/note="gene_id:K14A17.6"
/pseudo
/codon_start=1
/evidence=not_experimental
/product="liponamide dehydrogenase"
join(29096, .29206,29695, .29826,29925, .30009,30149, .30312,
30830, .31267)
/note="contains similarity to similar to ubiquitin
conjugating enzyme
gene_id:K14A17.7"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94978.1"
/db_xref="GI:7670024"
/translation="MADERYNRKNPAVKRIQEVKEMOANSDDFMSLPLEINIFEWQ
FAIRGPDGTEFEGIVGHRIQPADYPPKPPSMLLTPNGRETWTIKCLSLSNVHPE
HMOPSVNLTVALIAFMPTSPNGALGSVDYKDPKRRFLAKSRETPEPKYSPQRK
LIDEIHQYILSKATVYKPLPLECSOAPSIVSEANSOPEQAIIVTIRKSYLKFLAIPK
VDQIIEITAEAVNTASVVPAAAPLVAPEVSVKASVSGEQRMARRAAQKPPVDDRLFT
WAAGIETIAIMVLLLLKKFKISGYSTGFEMDDGS"
join(32260, .32364,33121, .33412,33506, .33750,33837, .34103)
/note="gb|AAD26962.1
gene_id:K14A17.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94979.1"
/db_xref="GI:7670025"
/translation="MGKSSNIVFDRIKEBROMLSFFKIFORADLSSDCMRALPYSFIS
KYSEKDFSVKMWIRAQWGTWDEVSKNPTVYIETRGWDLVPSNALQNEFFIFTH
RGNMVFHVNLYEQNGLEMKPKRFQTMGFSGIIKKEGNSLIDVKKKEESDESPRA
EFLVPRKKTEDSKSKMTRNKRKSKSKQVLDGVPEFKITIRKSYLKFLAIPK
HFVDDHIPNKSIFITIRHPNGGWSKVLCLVREITIFSGGYSKLAREPFLMVGDKCTF
KLKPFPEVLLTSKKNREKMDQCMID"
complement(join(34528, .34725,35213, .35322,35680, .35863))
/note="gb|AAF26101.1
gene_id:K14A17.9
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94980.1"
/db_xref="GI:7670026"
/translation="NAESGREGIVAVDFDCSKKALSWAIDNVVRDGHLLITIAH
DMNYEGEQMLNETVGSFPFNPSEFSDAAVMKKYALKPDAETLIDLVNTAARKRTIVV
KHIYMGDPREKICAAEQIPLSSVLNMRGLGGLKRMIMGVSNNHVNNAVACPVTYVK
AH"
complement(join(36309, .36416,36556, .36846,37250, .37324,
37500, .37634,38020, .38145,38221, .38374,38433, .38573,
38878, .38987,39134, .39281,39474, .39751,39993, .40262))
/note="gene_id:K14A17.10
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94981.1"
/db_xref="GI:7670027"
/translation="MGDTNGASLIEIGQEEVEFPFLAFLDYARTVISPEDEDEKEE

Query Match 33.9%; Score 576.2; DB 8; Length 92x20;
Best Local Similarity 99.5%; Pred. No. 1.1e-90;
Matches 578; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGGTTTTTATGGATAACATGACAAATATTTATTTATTTTCATGAGITTTATTTGATAG 60
|||||
Db 71476 TGGGTTTTTATGGATAACATGACAAATATTTATTTATTTTCATGAGITTTATTTGATAG 71535
|||||

QY 61 CATGACAAATATTAATATATCATGTTGTTAATAACATGTTTGTCTTATAATACATGCAAT 120
|||||
Db 71536 CATGACAAATATTAATATATCATGTTGTTAATAACATGTTTGTCTTATAATACATGCAAT 71595
|||||

QY 121 TTAATAATCAGACATTTGTTTTTAAATCAATCTAATCTCTTATATCAACAACGCAATTGAC 180
|||||
Db 71596 TTAATAATCAGACATTTGTTTTTAAATCAATCTAATCTCTTATATCAACAACGCAATTGAC 71655
|||||

QY 181 GGAAATTCAGGTAAAGAGAAATAAAGATGAGAGATAGATAGATTCTAAGAAA 240
|||||
Db 71656 GGAAATTCAGGTAAAGAGAAATAAAGATGAGAGATAGATAGATTCTAAGAAA 71715
|||||

QY 241 AGAAGAGAGAACATGTAGGTGAACAAAAATAAAGAGATATGATATATTTTATGAGAG 300
|||||
Db 71716 AGAAGAGAGAACATGTAGGTGAACAAAAATAAAGAGATATGATATATTTTATGAGAG 71775
|||||

QY 301 GTGGTGAAGATTATTTAGGAGAGGAGAGAGAAATAGAAAAAATAATGACATGGTGAA 360
|||||
Db 71776 GTGGTGAAGATTATTTAGGAGAGGAGAGAGAAATAGAAAAAATAATGACATGGTGAA 71835
|||||

QY 361 TCTGAAGAAGATGAATTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTC 420
|||||
Db 71836 TCTGAAGAAGATGAATTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTC 71895
|||||

QY 421 TCGTAAAGAAGATGAAAAAGAAACAAAAAGAGAAAGAAAGAAAGAGAAAGAGAAAGAGAA 480
|||||
Db 71896 TCGTAAAGAAGATGAAAAAGAAACAAAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 71955
|||||

QY 481 CTAACATATTCGCAAAATTTCTTACCGCACAATACTATTGTCACAGGTATTTTGTG 540
|||||
Db 71956 CTAACATATTCGCAAAATTTCTTACCGCACAATACTATTGTCACAGGTATTTTGTG 72015
|||||

QY 541 TATCTTTTGAAGTCAAAAGTTATTTTACATATACTCTA 581
|||||
Db 72016 TATCTTTTGAAGTCAAAAGTTATTTTACATATACTCTA 72056
|||||

RESULT 3
AB005248/c
LOCUS
AB005248.1
DEFINITION
Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MX110.
ACCESSION
AB005248.1
VERSION
AB005248.1
KEYWORDS
GI:2264320

Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
clone:MX110.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (sites)
Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, S., Fukami, M.,
Miya, J., N. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned pl clones
DNA Res. 4 (3), 215-230 (1997)
97471969
REFERENCE
2 (bases 1 to 83646)
Nakamura, Y.
Direct Submission
Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1-32-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:y.nakamura@kazusa.or.jp,


```
CDS
join(42817..43766,43839..43979,44083..44403)
/note="emb|CAB86928.1
gene_id:MX110.10
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
CDS
join(45089..45677,46843..47158,47312..47452,47544..47871)
/note="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:MX110.11"
/pseudo
/codon_start=1
/evidence=not_experimental
48978..49433
/note="gene_id:MX110.12
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09352.1"
/db_xref="GI:9758818"
/translation="MHVDSMKRARSVAVEGKERNRSSNGEETRRQSKHARTAKAR
MNYKESKPLPAPKKSLSMLISLRKEDTQDDNNVIVVEKEKEKEKGTIVONIRE
NAEMLSKCVVEPHMLTYEEWTFAFAGWGQWQYDAVMSGKSWFN"
complement(join(50633..50872,51043..51177,51268..51438))
/note="gene_id:MX110.13
sp|P10798"
/codon_start=1
/evidence=not_experimental
Query Match 31.8%; Score 540.2; DB 8; Length 83646;
Best Local Similarity 93.6%; Pred. No. 2e-84;
Matches 574; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1 TGGGTTTTTATGATACATGACAAATATTATTATTTTCATGAGTTTATTGGATAG 60
DB 23771 TGGGTTTTTATGATACATGACAAATATTATTATTTTCATGAGTTTATTGGATAG 23712
QY 61 CATGACAAATATTAATATATATCAGTCTTAATAACATGTTTCTTAAATACATGCAAT 120
DB 23711 CATGACAAATATTAATATATCAGTCTTAATAACATGTTTCTTAAATACATGCAAT 23652
QY 121 TTAATACAGACATTTGTTTAAATCAAAATCTAATCTCTATATATCAACAGCATGAC 180
DB 23651 TTAATACAGACATTTGTTTAAATCAAAATCTAATCTCTATATCAACAGCATGAC 23592
QY 181 GGAAATTCAGTAAAGAGAAATTAAGAAATGAGAGATAGAGAGATTCTATGGAAA 240
DB 23591 GGAAATTCAGTAAAGAGAAATTAAGAAATGAGAGATAGAGAGATTCTATGGAAA 23533
QY 241 AGAAGAGAGACATGTAGGTGAACAAAATAAGAGATATGATATATTTTATGAGAG 300
DB 23532 AGAAGAGAGACATGTGGGTGACAAAATAAGAGATATGATATATTTATGAAG 23473
QY 301 GTGTTGAAGATTATTTTAGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGTTGAA 360
DB 23472 GTGGTGAAGATTATTTTAGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGTTGAA 23413
QY 361 TCTGAAGAGATGAATGTGTTGAAGATGAAGAGAGAAAGAGAACTCCATGGGTAAAGTC 420
DB 23412 TCTGAAGAGATGAATGTGTTGAAGATGAAGAGAGAAAGAGAACTCTATGACTAAAGTC 23353
QY 421 TCGTAAAGACATGAAAAGAAACAAAGAGGAGGAGAGAAAGAGAGAGAGCTAAATAGA 480
DB 23352 TCGTAAAGACATGAAAAGAAACAAAGAGGAGGAGGAGGAGAGAGAGCTAAATAGA 23293
QY 481 CTAACATATGCCAAATCTCTGAGCCGACAAATFACATTTTGGTCCAAAGTTATTGTTG 540
DB 23292 CTAACATATGCCAAATCTCTGAGCCGACAAATFACATTTTGGTCCAAAGTTATTGTTG 23233
QY 541 TATCTCTTTGAAGTCAAAAGTTATTTCTTACATATATCTCTAAATATATACCGATACCAA 600
DB 23232 TATCTCTTTGAAGTCAAAAGTTATTTCTTACATATATCTCTAAATATATACCGATACCAA 23173
```

```
QY 601 TTTTTCACACAT 613
DB 23172 TTATTTAAATAT 23160

RESULT 4
AC012394 94487 bp DNA linear PLN 12-OCT-2000
LOCUS Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence,
DEFINITION complete sequence.
AC012394
AC012394.3 GI:6554469
HTG
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 94487)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence
Unpublished
REFERENCE 2 (bases 1 to 94487)
AUTHORS Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL
REFERENCE 3 (bases 1 to 94487)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 4 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 5 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE 6 (bases 1 to 94487)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT On Dec 10, 1999 this sequence version replaced gi:6143858.
Address all correspondence to:at@tigr.org
```

BAC clone F15M4 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from Sp6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, <http://www.tigr.org/soflab/glimmerm.htm>), and Geneslicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tqi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by


```

tRNAScan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAScan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
    source                1..94487
                        /organism="Arabidopsis thaliana"
                        /cultivar="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="1"
                        /clone="F15M4"
                        complement(967..1017)
                        /rpt_family="AT_rich"
                        complement(1289..1338)
                        /rpt_family="AT_rich"
                        complement(4053..4080)
                        /rpt_family="AT_rich"
                        complement(6943..9510)
                        /gene="F15M4.1"
                        /note="identical to hydroxy methylglutaryl CoA reductase
                        (AA 1-592) GB:X15032 [Arabidopsis thaliana]"
                        complement(join(<6943..7515,7607..7953,8115..8296,
                        8522..>9510))
                        /gene="F15M4.1"
                        complement(join(7255..7515,7607..7953,8115..8296,
                        8522..9510))
                        /gene="F15M4.1"
                        /codon_start=1
                        /product="hydroxy methylglutaryl CoA reductase (AA 1-592);
                        9510-7255"
                        /protein_id="AAF16652.1"
                        /db_xref="GI:6554470"
                        /translation="MDLRPPPPVPTNNNSGSRSPQRTSDDHRRRTATTIAPP
                        PKASDALPLPLNTNAVFETLFSSVAYILLHRWDKIRYNTPLHVVTITELGAIILI
                        ASYILLGFGDFQVFSISASGDADLDATDDDDHRLVTCSPPTPIVSVAKLPNP
                        EPITVESDEDEIKVSDVIGVIPSISLESRLGDCCKRAASIRREALQVTRGSIEGL
                        PLGDFYESILGQCCMPGVYIPIPVGIAGPLLLDGDYEVSPMATTEGCLVASTNRGC
                        KAMFISGATSTVLDKMDTRAPVVRFAASRASSELKFELENNPENTFLAVWNRSSRF
                        ARLOSQKTAAGNVVRFCCSTGDAMGNMVKQNVLEYLTDLPDMDVIGISGN
                        FCSQKPAANWTEGKRSVVCVAVIRGEIVNKVLKTSVAALVELNMLNLAGSAVAG
                        SLGGFNASHNIVAFIANGQPAQNVSSQCITMMEAINDGDKDITHSVTPMSIEVG
                        TVGGTQLASQSCDLNLGKVGASTSPGNRRRLATIVAGAVLAGELSLSMSAIAGQ
                        LVRSRMKRSRSDISGATTTTTTT"
                        complement(7258..7283)
                        /rpt_family="(CAA)n"
                        complement(11154..11209)
                        /rpt_family="AT_rich"
                        complement(11551..11580)
                        /rpt_family="AT_rich"
                        complement(11933..12004)
                        /rpt_family="AT_rich"
                        12390..13268
                        /gene="F15M4.2"
                        /note="similar to unknown protein GB:AAD30603 [Arabidopsis
                        thaliana]"
                        join(<12390..12391,12491..12629,12730..12861,12929..13019,
                        13099..>13268)
                        /gene="F15M4.2"
                        join(12390..12391,12491..12629,12730..12861,12929..13019,
                        13099..13268)
                        /gene="F15M4.2"
                        /codon_start=1
                        /product="unknown protein; 12390-13268"
                        /protein_id="AAF16653.1"
                        /db_xref="GI:6554471"
                        /translation="MFQDKALYCKELWRRETLPLQSGSRVVKLQGLKSNWSVEKISY
                        PASIPARVSQLLKNHMEKLNOMRRLLNTEKLIKFAESLKEVNNKAGLNVLVLEP
                        EGIIVAPNSRERYFIINYIVCEQMGIPYSSVVLVALCLVSVFIVPRSLPSSL
                        TKDQGLRSRSHRSKDS"
                        14056..15506
                        /gene="F15M4.3"
                        /note="similar to cinnamoyl-CoA reductase GB:CAA56103
                        [Eucalyptus gunnii]"
                        join(<14056..14170,14277..14431,14516..14701,14819..14990,

```

```

15096..15285,15362..>15506)
/gene="F15M4.3"
join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)
/gene="F15M4.3"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 14056-15506"
/protein_id="AAF16654.1"
/db_xref="GI:6554472"
/translation="MAYKQKVCVTGAGGFASWVKFLSRGYTVHCTVRDPCDEKND
HLRLDNASKNLKLFKADLDFDEGLFSAIDGCSGVFLASPVPEGSELKLPALTGPK
NVLEACTETKVQKVYVSSIAAVYNPKVQVAKEDCWSDFQILHSLSGYWRYYIL
AKLTREALEWSKRNFAFVVTCLPQSVIIGRPLQSVILNLSGLLKKLKGIGKLSLSD
ELYLDVRDVAADALLVKNREATGRYICNSFSYTLSEKLNGLMPPYKRNFPESFTE
VKEKEVRPLSAEKLKNLGKFRPLEETIDDSVVVSFEAAGDLPA"
16281..16361
/rpt_family="(CAT)n"
16955..18423
/gene="F15M4.4"
/note="contains Pfam profile: PF00C76 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
join(<16955..17161,17385..17529,17633..17755,17844..17884,
17978..18215,18332..>18423)
/gene="F15M4.4"
join(16955..17161,17385..17529,17633..17755,17844..17884,
17978..18215,18332..18423)
/gene="F15M4.4"
/codon_start=1
/product="putative RNA-binding protein; 16955-18423"
/protein_id="AAF16655.1"
/db_xref="GI:6554473"
/translation="MAYQVPVSGSFHYNSPFGDITFKVVFVGLAMETQSETLRQHF
EQYGLLEAVYIADKNTGRSGYGVTFRDEPAARACACADPTIIDGRRANCNLASLG
RPPPLPVAIVPNNPDLPTLEMCRLVLEHLEIRLSNFHFTTTKSLSLMTAY
GPEYMSQSGLYSPYMQQYLOVYVPGAVNSPYQYQGLSQTIPNGHGYTAVQGS
VPGSHILQGLPTVSTWTSSMPALQAPYPSGICPEKAVQSHIIVHSPQFMQSTASDQ
TTR"
complement(19352..20843)
/gene="F15M4.5"
complement(join(<19352..19378,19453..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))
/gene="F15M4.5"
complement(join(19352..19378,19453..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))
/gene="F15M4.5"
/codon_start=1
/product="unknown protein; 20843-19352"
/protein_id="AAF16656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSSPQSFNSPRVTITDSRRCSISAISLVLDSSNEEQ
HRISSRDHVGRKRDVMLQIASVFFELPLAISPAFAETNASEAFVYVDETNKFEISI
POVSIATIGLGFDPFDMESFGKVEAFETLVSGLDPSWPKPVGTAKLIDRSRSGEY
VLEYTLQNPGEARKHLYSAIGMATNGWYNRLYTVTQFTDESAEQSSKIQKTVKSR
FI"
complement(20998..21026)
/rpt_family="AT_rich"
complement(21140..21167)
/rpt_family="AT_rich"
22462..23090
/gene="F15M4.6"

Query Match          31.4%; Score 534.6; DB 8; Length 94487;
Best Local Similarity 96.5%; Pred. NO. 1.8e-83;
Matches 557; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

QY 1 TGGGTTTTTATTCGATAACATGACAATATTTATTTATTCATGACTTTTATTCGATAG 60
|||||
Db 1917 TGGGTTTTTATTCGATAACATGACAATATTTATTTATTCATGACTTTTATTCGATAG 1976
|||||

QY 61 CATGACAAATATTAATATATCATCTGTTAATAACATCTTTTGTTCCTTAAATACATCAATT 120
|||||

```



```
Db 1977 CATGCAAAATTAATATATATTAGTGTCTTAATAACATGTTTTTGTCTTAAATACATGCATT 2036
QY 121 TTAATAATCAGACATTTGTTTTTAAATCAATCTTAATCTCTTATATACAAACGACATTTGAC 180
Db 2037 TTAATAATCAGATATTGTTTTTAAATCAATTTAATCTCTTATATACAAACGACATTTGAC 2096
QY 181 GGAAATTCAGTAAAAAGAGAGAAAAATAAGAAATGAGATAGAGAGATTTCTATGGAAAA 240
Db 2097 GGAGAAATTCAGCTAAAAAGAGAGAAAAATAAGAAATGAGATAGAGAGATTTCTATGGAAAA 2156
QY 241 AGAAAGAGAGACATCTAGGTGAAC-AAAATAAGAGATATGATGATATATTTATGAGA 299
Db 2157 AGAAAGAGAGACATCTGGGTGAACAAAATAAGAGATATGATGATATATTTATGAGA 2216
QY 300 GGTGTGAAGATTATTTAGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGA 359
Db 2217 GGTGTGAAGATTATTTAGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGA 2276
QY 360 ATCTGAAGAGATGAATTTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGT 419
Db 2277 ATCTGAAGAGATGAATTTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGT 2336
QY 420 CTCGTAAGAGATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
Db 2337 CTCGTAAGAGATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2396
QY 480 ACTAACTATTGCCAAATTTCTGTAGCCGACAAATACTATTGGTCCAAAGGTTATTTTGT 539
Db 2397 ACTAAATATTGCCAAATTTCTGTAGCCGACAAAGTACTATTGGTCTTAAGTTATTTGT 2456
QY 540 GTATTCTTTTGAAGTCAAAAGTATTTCTTTACATATA 576
Db 2457 GTATTCTTTTCAACTCAAAAGGTTATTTCTTACATATA 2493

RESULT 5
LOCUS AC015450/c
DEFINITION Arabidopsis thaliana chromosome 1 BAC F1466 genomic sequence,
complete sequence.
ACCESSION AC015450
VERSION AC015450.5 GI:12323968
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 100806)
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F1466 genomic sequence
Unpublished
2 (bases 1 to 100806)
Lin,X. and Kaul,S.
Direct Submission
Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 100806)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280761.
Address all correspondence to:at@tigr.org
```

BAC clone F1466 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,

http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tldb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smith,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

source

```
1. .100806
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F1466"
```

misc_feature

```
1
/note="35000 nt before this point were not included in  
the submitted sequence, due to overlap with another BAC  
[F15M4]"
```

gene

```
/gene="F1466.1"
/note="contains Pfam profile: PF00097 Zinc finger, C3HC4  
type (RING finger)"
```

mRNA

```
<5217..>5774
/gene="F1466.1"
```

CDS

```
5217..5774
/gene="F1466.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
```

```
/protein_id="AAG51946.1"
/db_xref="GI:12323975"
/translation="MARLLRLQEAANSTPAEASPPFNSDLVLILAVLALCTIIG  
LIAVRCAMLRRIASNRSDQTHPPVAANKGLKKVLRSLPKLYSPDPAPAEKLV  
ECAICLTFEAGDELRLVPCQGHGFHVCIDTWLGSWSCPSCRQLIVVTRCHRCGL  
PGSSSSGPEPTRIKQREDGPDNLP"
```

gene

```
/gene="F1466.2"
/note="N-term similar to N-term of NAM GB:CAA63101  
[Petunia x hybrida] (apical meristem formation), CUC2  
GB:BAA19529 [Arabidopsis thaliana], GRAB2 protein  
GB:CAA09372 [Triticum sp.]"
```

mRNA

```
complement(join(<8331..8855,9526..9800,9933..>10137))  
/gene="F1466.2"
```

CDS

```
complement(join(8331..8855,9526..9800,9933..>10137))  
/gene="F1466.2"  
/codon_start=1  
/product="unknown protein; 10137-8331"  
/protein_id="AAG51953.1"  
/db_xref="GI:12323982"  
/translation="MMLAVDLSELAGEERNRGLPFFRHFPTDELIITFYLSAKI  
FHGLSGIHISEVDNLNCEPWEPMKMERWIFYSLDRKYPTGLRNRATAGY  
WKATGKDEVFSGGGQLVGMKKTLYFKRAPRLKTKWVMEYRLNHRHTCK  
BEWICRVFNKTKGDKVGLHNOISYLHNHSLSTTHHHHEALPLIEPSNKTLTNF  
PSLLYDDPHQNNNNFHLGSSGHNIDELKALIN2VVSOLNGIIFPSGNNDDEDD  
ENLGVKTFQSSNGNEIDVRDYLENPLQEASVGLGFSSSPGPLMLLDSPCLGFLQ  
"
```

gene

```
complement(15417..18176)  
/gene="F1466.3"  
/note="Similar to phosphate transporter GB:BAA20522  
[Catharanthus roseus]; contains Pfam profile: PF00083  
Sugar (and other) transporter"
```

mRNA

```
complement(join(<15417..16305,17467..>18176))  
/gene="F1466.3"
```

CDS

```
complement(join(15417..16305,17467..18176))
```

mRNA	/gene="F1466.3"		mRNA	complement(join(<26257..26401,26478..26667,26773..26935,27062..27247,27332..27486,27593..>27736))	
	/codon_start=1			/gene="F1466.7"	
	/product="putative phosphate transporter; 18176-15417"			/note="contains nonconsensus splice site"	
	/protein_id="AAG51941.1"			complement(join(26257..26401,26478..26667,26773..26935,27062..27247,27332..27486,27593..27707))	
	/db_xref="GI:12323970"			/gene="F1466.7"	
CDS	/translation="MPELSLSALDARIQWYHFKALIVAGMGLFTDYLDFLCIAPIMKMSIQYVHKDSITGALLSTSYAIALIGTALGOLIFGYLGDVRGRKRVGLSLIMVFFSSRCGSCVCTRSQVMSVSGFERVLGIGDYPLSATIMSEFANKRTRGAFIAAVFMQGLGILMSNVTHVCLAPKNAGESSKTNVAGLETFLAPESDIARLLIIGALPAALFYWRMLMPETARYTALVENVNVQAKDMQVMSVMSIQITEDSSSELEQPPSSSYKLFRRRLSLGRDLFAASANFLVDVFTSNLLLSQIFNSQIMLNSTNVYDSAFYAKLAAIVAACTIPGYWFTYVIFDKIGRKIQMMGFFLMAVLYLVAGIPYSWYKSHKTKNKGVLVGLIFFNSFGPNTTTFIPAEFPARFSTCHGISGAAGKF GAIVGTGFWLATRHHEEDGFPDVKRVRIAFILILGVCYACIMIVTVLFTRETMGRSLE ENEDIYSTSGSPANELLRROY"		CDS	/note="contains nonconsensus splice site"	
	/codon_start=1			/codon_start=1	
	/product="hypothetical protein; 19301-18673"			/product="putative cinnamoyl-CoA reductase; 27707-26257"	
	/protein_id="AAG51943.1"			/protein_id="AAG51951.1"	
	/db_xref="GI:12323972"			/db_xref="GI:12323980"	
gene	/translation="MNAENQTTTHSKVISHVCTGAKLGSVGPPIGLVDIGVSEVAYIFRVLPGLEKQNDKIKCEIQREGVCIQGVPIEAIIPSDTGCLYRMQVQQLCPPG PFSITFNLPGVDPLRFLSPNFRSDGIFEVVVKLGVRIPIS"		gene	/translation="MAVKQKVCVTGAGGFIASLWYKFLLSRGYTVHGTVRPPDCDKNDHURKLDNASKNLKFLKADLFDEGLFSAIDGSGVFHIASPVPFEGBELIKPALFTGK NVLEACTFKYKQVWVSSIAAAYVNPKPQDVAKDEDCWSDTQYLSHLSRYRYLAKT LTERALEWSKRFENADVTCLPSVLIIGPRLQSLNSSLGLLKF1KGIGIKLSLSDEL YLVDRVDADALLLYVENREATGRYICNSHLSYTDSDIMEK1KNMYPKRNPFESFTEVKE KVRPLSAEKLNLGWKFRPLEET1DDSVSVSEAAAGDLPKA"	
	/gene="F1466.5"			/gene="F1466.7"	
	join(<20920..20986,21073..21148,21227..21330,21407..21472,21739..21821,21911..21986,22064..22182,22264..22308,22385..22411)			Query Match 31.4%; Score 534.6; DB 8; Length 100806;	
	/gene="F1466.5"			Best Local Similarity 96.5%; Pred. No. 1.8e-83;	
	join(20920..20986,21073..21148,21227..21330,21407..21472,21739..21821,21911..21986,22064..22182,22264..22308,22385..22411)			Matches 557; Conservative 0; Mismatches 19; Indels 1; Gaps 1;	
CDS	/codon_start=1		CDS	QY 1 TGGGTTTTTATTGGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 60	
	/product="unknown protein; 20920-22411"			Db 39846 TGGGTTTTTATTGGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 39787	
	/protein_id="AAG51945.1"			QY 61 CATGACAAATATTAATATATCATGTTGTTAATAACATGTTTTTGTTCCTTAAATAACATGCATT 120	
	/db_xref="GI:12323974"			Db 39786 CATGACAAATATTAATATATAGTGTTAATAACATGTTTTTGTTCCTTAAATAACATGCATT 39727	
	/translation="MAAISPLWSSPQSFNSPRVITDTSRRSSISAISVLDSSNEQHRISRDHVKRRDVLQIASVFFFLPIAISPAFETNASEAFRVYTDENKFEISI PQVSTAITGLGDFTRMESGKVEAFATLVSLDRSNQKPVGVYAKLIDRSKGFY YIEYTLQNGEARKHLYSAIGNATNGWYNRLYTVTGQFTDESAEQSSKIQKTVKSFR FI"			QY 121 TTAATAATCAGACATTTGTTTTTAAAAATCAATCTAATCTCTTATATCACAACGACATTGAC 180	
gene	complement(23340..24808)		gene	Db 39726 TTAATAATCAGATATTTGTTTTAAAAATCAAAATTTAATCTCTTATATCACAACGACATTGAC 39667	
	/gene="F1466.6"			QY 181 GGRAAAATTCAGGTAAGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGATTTCTATGGAATAA 240	
	/note="contains Pfam profile: PF00076 RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)"			Db 39666 GGAGAAATTCAGGTAAGGTAAGAGAGAGAGAGAGAGAGAGAGAGATTTCTATGGAATAA 39607	
	complement(join(<23340..23431,23548..23785,23879..23919,24008..24124,24234..24378,24602..>24808))			QY 241 AGAAAGAGAGAACATGTAGGTGAAC-AAAAATAAGAGATATGATGATATATTTTATGAGA 299	
	/gene="F1466.6"			Db 39606 AGAAAGAGAGAACATGTGGGTGAACAAAAATAAGAGATATGATGATATATTTTATGAGA 39547	
mRNA	/translation="MAAISPLWSSPQSFNSPRVITDTSRRSSISAISVLDSSNEQHRISRDHVKRRDVLQIASVFFFLPIAISPAFETNASEAFRVYTDENKFEISI PQVSTAITGLGDFTRMESGKVEAFATLVSLDRSNQKPVGVYAKLIDRSKGFY YIEYTLQNGEARKHLYSAIGNATNGWYNRLYTVTGQFTDESAEQSSKIQKTVKSFR FI"		mRNA	QY 300 GCTGCTCAAGATTTATTTAGCAGAGCGCAG	

RESULT 6
 ATT32A17/c Arabidopsis thaliana DNA linear PLN 23-MAR-2001
 LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T32A17 (ESSA project).
 DEFINITION
 accession AL161813 AC007635
 version AL161813.1 GI:7321058
 keywords
 source thale cress.

		MQQLEEVAMGGIPPIPCGGRIILDIISERDGDGKGRFYECTNWKNDGLHIRKLWDKA MYEVRNLREQVDNHHQKIQSLSEYINQOMRSKLUDEIQKKMGTLLWRICCGE"
exon		complement(28303. .28494) /gene="AT4g08720" /number=1 complement(28495. .28564) /gene="AT4g08720" /number=1 complement(28565. .28761) /gene="AT4g08720" /number=2 complement(28762. .29625) /gene="AT4g08720" /number=2 complement(29626. .30162) /gene="AT4g08720" /number=3 complement(30163. .30332) /gene="AT4g08720" /number=3 complement(30333. .31953) /gene="AT4g08720" /number=4 34618. .36007 /gene="AT4g08730" complement(join(34618. .34852,35539. .35690,35828. .36007)) /gene="AT4g08730" complement(join(34618. .34852,35539. .35690,35828. .36007)) /gene="AT4g08730" /codon_start=1 /product="hypothetical protein" /protein_id="CAB82109.1" /db_xref="GI:7321062" /translation="MEGTFNLOVMETKVDLLARKVDLLTSKMDLCATNVDLLTSKNN LCATSDKAMLTSSKGNKVPKAKTLAASDSESSSSSDSDIEELCFRFLIKIVQ LRPSYFIIFADIDKEGANKALELNGSDMAGKELVVATALLMRDSYLGYSNCGGIGG RFGGRFSGVGIGFGRGCGGLGQYYAQV" complement(34618. .34852) /gene="AT4g08730" /number=1 complement(34853. .35538) /gene="AT4g08730" /number=1 complement(35539. .35690) /gene="AT4g08730" /number=2 complement(35691. .35827)
intron		
Query Match	29.4%;	Score 499.8; DB 8; Length 104386;
Best Local Similarity	90.6%;	Pred. No.1.9e-77;
Matches 557; Conservative	0;	Mismatches 52; Indels 6; Gaps 2;
Qy	1	TGGGTTTTTATGGATAACATGACAAATATTTTATTTATTTTCATGAGCTTTTATTTGGATAG 60
Db	87736	TCGGTTTTTATGGATAACATGACTATATTTATATATTTTCATGAGCTTTTGTGGATAG 87677
Qy	61	CATGACAAATATTAATATATCAGTGTGTAAPACATGTTTGGTTCTTTAAATACATGCATT 120
Db	87676	CATGACAAATATTAATATATCAGTGTGTAAPACATGTTTGGTTCTTTAAATACATGCATT 87617
Qy	121	TTAAATTCAGACATTTGTTTTTAAATCAAAATCTAATCTCTATATCACACGACATTGAC 180
Db	87616	TTAAATTCAGACATTTGTTTTTACATCAAAATCTAATCTCTATATCACACGACATTGAT 87557
Qy	181	GGAAATTCAGGTAAAAAGAGAAAAATAAAGATGAGAGATAGAGAGATTTCTATGAAAAA 240
Db	87556	GGAGATTCAGGCAAAAAGAGAAAAACAAGATGAGAGATAGAGATTTCTATGAAAAA 87497
Qy	241	AGAAAGAGAGACATGTAGTGTGACAAAAATAAGAGATATGATGATATTTTTATGAGAG 300
Db	87496	ATAAAGAGGGAACATGTGGGTGTAAACAAATAAAGAGATATGATGATATTTTTATGAGAG 87437
Qy	301	GTGCTGAAGATTATTT-----TAGGAGAGGGAGAGAGAAATAGAAAAAGAAATGACATG 355


```

/misc_feature
    /chromosome="iv"
    /map="18.8 cM"
    /clone="BAC F9H3"
    /note="hybridizing YAC is C1C8B1"
    1..31783
    /notes="overlap with T5L23, GenBank accession number
    AC005142, from position 26 to 31808"
    851..1719
    /gene="F9H3.20"
    /pseudo
    /evidence=not_experimental
    complement(981..1120)
    /note="function-unclassified; similar to F6N23, GenBank
    accession number AF058919"
    /rpt_type=dispersed
    3173..3969
    /note="function-putative_transposon; similar to FCAl,
    GenBank accession number Z97336; moderately similar to
    reverse transcriptase gene of DW15 non-LTR
    retrotransposon, GenBank accession number L47193"
    /rpt_type=dispersed
    5889..6303
    /note="function-chloroplast_repeat"
    /rpt_type=dispersed
    complement(6533..12091)
    /gene="F9H3.19"
    /note="encodes putative calcium channel; genomic copy of
    EST W43766; genomic copy of EST T43251; EST T43251
    apparently encodes intronic sequence; possible alternate
    exon 1 from position 12772 to 12846 is predicted by a
    NetPlantGene splice donor but has no EST coverage; gene
    model last edited on 27 Jan 99"
    /evidence=not_experimental
    complement(join(6653..6746,6833..6870,6991..7062,
    7175..7261,7436..7526,7682..7746,7985..8153,8229..8292,
    8380..8487,8757..8814,8891..9034,9127..9192,9311..9383,
    9530..9591,9745..9822,9929..10034,10127..10293,
    10391..10451,10557..10580,10644..10774,10890..10930,
    11292..11403,11500..11643,11972..12091))
    /gene="F9H3.19"
    /note="functional catalog ID=07.01"
    /codon_start=1
    /evidence=not_experimental
    /product="putative calcium channel"
    /protein_id="AAD11598.1"
    /db_xref="GI:4206210"
    /translation="MEDPLGRDLSGGGTRDRRSEAIHTGTPPQKAALVDLAEDG
    IGLVEILDQSFGESARYIFTRLDLIWSLNYFALLFLNFFQPLNCKNPKPSCK
    DRDYLLGELPYLTNAESIIEASSRIEFTSRLNLVKVAVILFVDVLDVLYLSPL
    AFDFLPFIAPVRVYIFILSIQRQCADHVMELRDLVLLSGMLGTYLNLALWMLFL
    LPASAFVMEEDTQGLTVFTSYGATLYQMFIPTTSNPDVMTIPAKKSRWSSVFE
    VLYLVGVVFTNLILAVVDSFEKQLAKQVSGDMQKRMLEKAFGLIDSQKGEID
    KQCKILKEQLTNYRTPKISKEEFLGFDELDTRDFKINKEDEFADICQAIARFQK
    EBPVSLFPHFOIYHSALQRAFRVSPNFGYAFISFLINFIATVAVVETLDEESS
    AOKPQVAFVFGVYVLEMAKIKYTYFENYREGANRDFLVTWVIVIGETATFIT
    PDENFTFNGEMIRYLLRLMLRILRLNMVQRVRAFTFITLIPSPVLYGTIFCV
    LCYCSIGVQVFGGVNAGNKKLPETELAEDDYLLFNENDYPNGMVTLLNLYMGNQ
    VMKESYKDLCTGWSYITVFSFYVITILLNLNVVAFVLEAFFTELDEEEKCGQGD
    SQEKNRRRSAGSKRSORVDLTHHMLGDELSEKPECSSTDT"
    complement(13149..18831)
    /gene="F9H3.18"
    /note="encodes putative glucan synthase component; gene
    model last edited on 5 Jan 99"
    /evidence=not_experimental
    complement(join(13149..14597,14778..16800,16961..18831))
    /gene="F9H3.18"
    /note="similar to 1,3-beta glucan synthase; functional
    catalog ID=01.05.99; functional catalog ID=09.01"
    /codon_start=1
    /evidence=not_experimental
    /product="putative glucan synthase component"
    /protein_id="AAD11597.1"
    /db_xref="GI:4206209"

/translation="MSLRHRTVPQTPGRPLAAEAVGIEEPEYNIIPVNNLLADHPSLR
FPEVRAAAALKTVDGURRPYPVOMRSHYDLDLWALFLFFQKDNVNRQREHVMVLHLA
NAQRLSPDPNIDSLDSAVVRRFRKLLANYSSW:SYLCKKNIMISDRNPSRREL
LYGLYLLIMGEANRLMEPECTCYIFHNASELNKILEDCLENDTQGPYLPISLGEN
LFTGVVKPYDITQAEITDSKTSVAHCKNRVDDINEYFWTDRCFSKLWPLDLGS
NFFKSRGKSVGKTGFVERRTFFLYRSDFDRLWMLALFOAAIIIAWEKRPDTSVTR
OLWNAKARDVQVRLTLVLTWSGRMLLOALDAASQYPLVSRTEKHFHFMKLVIA
AANWIVAFVLYINWIKQKQDQKQNSAAITFIYOFIYAGAFALVPELLALALFIIPW
MRNFEETNMKIFLFTWFGQKSFVGRGLRGLVDNIKYKSTFIWFIATLFTFSYFL
QVKPMIPKSKLLWNLDVDEWHQFQGSNRI:SVAILMPELVLIYIMDIQIYAIYSS
IVGAVVGLFDHGLGIRDGQLRLRFQFASAIQNLMPPEQLLNARFGNFKDGLHR
LKLYRGRPFKLESNOVEANKFALINWEI:LAFREEDIVSDREVEALLPKNSWDV
TIWPCFLCNELLALSOARELIDAPDKW:WHKICKNEVRCACVAVDSIKHLL
SIJKVDTEHSILTVFQIINOSIQSOFTK:PRVDLLPKIYETLQKLGLVNDDETD
SGRVNVLOSLEYIETAFIEKTEITQELNSLGLTPSPASKLFLQNAIRKUPASND
FYRVRRLHILTSRDSMHSVPVNLAEARRIAFNSLFHNMHAPHOPEKMAFVSIT
PYSEEVYISKEQLRNETEDISTLYLYQTYADEVNKFERHREHREGIKTDSBLWTK
LRDLRLWASYGQTLARTVRGMYYRALKMLAFDLSAEMDIRAQEILGYSQKALQ
ELGQSDGFVSENDRLSSASSTLYKGRYGAALMKFTYVACQIYGSQAKKE
POAEIILYLMKONEALRIAYDEVPAGRGETUYSLVKYDHOLEKEVEIFRVKLPQ
VKLGECPENQNHAMIFTRGDAVQTDIMNQDSFEALAKMRNLLOEYNNHYHGRKPTI
LGVREHIFTGVSVAWPMASQETSFTLQGFVLANPLKVRMHYGHQPDVDFWFLSR
GGISKASVINISIEDIFAGNCTLRGNVTHFETIQVGRDVLGNQISMPFEAKVASG
NGEQLSRDVRRLGRLDFRMLFFSYFTVGI:FNTHVILTVYAFIWRVYLALSGV
EKSALADSTOTNALGVLNNOQFIQLGLFTALPM:VWLSLEEGFLLAIWNFTRMQIQ
LSAVYTFSMGTRAHVFGRTLHGGAKYRAT:GCVVVEHKGFTEFNRYLARSFKVRAI
ELGLLIVYASHSP:IAKDSLIYIANTITSNF:VISIMAPFVNPSPGFWLKVTVDFE
DFMMWITQGRISSTKSEKWEKWEQDHLNIGKAGLEVEILLVLRFFFOIGIYV
OLKANGSTSLFVLYFIYFAIFVLFLVICIYARDKYSKAHRYLQVFLVLI:LA
LVIIVALLEFTHSFIDFTLSLAPFTGWGII:IAQTORKWLKNTIFWNAVYVARM
YDILFGLIMVPAFLSMPGFSQMQTRILENAFSRGLRIMQIVTGCKSKGDV"
complement(20861..22302)
/gene="F9H3.17"
/note="encodes hypothetical protein; gene model last
edited on 5 Jan 99"
/evidence=not_experimental
complement(join(20861..21094,21437..22039,22145..22302))
/gene="F9H3.17"
/note="similar to F21B7.5, GenBank accession number
AC002560; functional catalog ID=19"

Query Match 29.2%; Score 497; DB 8; length 95519;
Best Local Similarity 92.5%; Pred. No. 5.8e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

Qy 1 TGGGTTTTTATGGATACATGACAAATATATTATTATTTTCATAGTTTATTGGATAG 60
Db 23303 TGAGTTTTTATGGATACATGACAAATATATTATATATATATATATTTTGTGGATAG 23362

Qy 61 CATGACAAATATTAATATATATATATAGTGTAAATACATGTTTCTTCTTAAATACATGCATT 120
Db 23363 CATGACAAATATTAATATATATATATAGTGTAAATACATGTTTCTTCTTAAATACATGCATT 23422

Qy 121 TTAATATCAGACATTGTTGTTTAAATCAATCTAATCTTATTCACACGACATTCGAC 180
Db 23423 TTAATATCAGACATTGTTGTTTAAATCAATCTAATCTTATTCACACGACATTCGAT 23482

Qy 181 GGAAATTCAGGTAAAAGAGAAATAAAGAAATGAGATGAGAGAGATTTCCTATGGAATA 240
Db 23483 GAAGAAATTCAGGCAAAAAGAGAAATAAAGAAATGAGAGATGAGAGATTTCCTATGGAATA 23542

Qy 241 AGAAGAGAGACATGTTAGGTGAACAAATAAAGAGATATGATATATATTTATGAGAG 300
Db 23543 AGAAGAGAGAACATGTTGGGTGAACAAATAAAGAGATATGATATATATTTATGAGAG 23602

Qy 301 GTGGTGAAGATTATTTAGGAGAGGGAGAGAAATAGAAAGAAATACACATGTTGTA 360
Db 23603 GTGGTGAAGA-TATTTAGGAGAGGGAGAGAAATAGAGAAAGAAATACACATGTTGTA 23661

Qy 361 TCTGAAGAAGATGAATTTGTGTTAAAGATGAGAGAGAAAGAGAACTCCATCGCTAAAGTC 420
Db 23662 TTTGAAGAAGATGAATTTGTGTTAAAGATGAGAGAGAAAGAGAACTTCATGGCTAAAGTT 23721

```


lemcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with ATCHRIV8 at the 5' end and an
 overlap with ATCHRIV10 at the 3' end.

FEATURES

Location/Qualifiers

source

1..159629

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

6741..9588

/gene="At4g03480"

complement(join(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588))

/gene="At4g03480"

complement(join(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588))

/gene="At4g03480"

/note="contains similarity to mammalian ankyrins
 similarity to similar to A. thaliana hypothetical protein
 F21B7.8, GenBank accession number AC002560
 Contains ATP/GTP-binding site motif A (P-loop) AA297-304
 contains EST gb:H37081, T42092"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB77833.1"

/db_xref="GI:7270671"

/translation="MVRSHLLRATPEASLPKSPSNCRSLSPVELRLYFIRODDF
 KKVDFIKTSYQNSLPCCDSEFLTSLRSDLENLPGEYVPMPEIFSAMRAGNI
 ELLEKLSYFTPMACKSDGDSVLHAAAGHLELVKNIITECPCLLEPKSKYQIP
 LHVAAAGRSAYKALVSLVSPRPVEDRDLNIIYVKIDGCTPLHAALKDLHE
 KAERIRKLSLHMHWRRCISFSDASTROMETAACLVNADQHASFLANKDGTSP
 LYLAVERNYSVLRAMLRPNKIQGKTSLASQLEGRSLHAAALKAKNTDVLNVIL
 NDDPSLVNERDEGRCTCLSGMSGYKIGIKLLDRSTKSVYECDDGSGFFIHMAVEK
 GHLKVVKEILKRPCDSKELVKNQGNMLHIAAKSAKVGSGFLGYIRRLDTENHLIEQ
 DVDGNAFLHATINWRCRTVDKLAFASTETKTLNTQNKDGLRPLDIAELNLPQDVL
 RERLTLMVLCVAPKSVGWLPVSGMTLSRSEPLDAKKYKOHINALLIVATIVATV
 FAAGFTIPGFSNAPNMGWATLADDSSTLFFVLDTLAMQSSIVAIVALI"

complement(6741..7064)

/gene="At4g03480"

/number=1

complement(7065..7160)

/gene="At4g03480"

/number=1

complement(7161..7708)

/gene="At4g03480"

/number=2

complement(7709..7782)

/gene="At4g03480"

/number=2

complement(7783..8101)

/gene="At4g03480"

/number=3

complement(8102..8408)

/gene="At4g03480"

/number=3

complement(8409..8866)

/gene="At4g03480"

/number=4

complement(8867..9282)

/gene="At4g03480"

/number=4

complement(9283..9343)

/gene="At4g03480"

/number=5

complement(9344..9444)

/gene="At4g03480"

/number=5

exon

complement(9445..9588)

/gene="At4g03480"

/number=6

10335..13501

/gene="At4g03490"

gene

complement(join(10335..10525,10983..11078,11590..12133,
 12345..12500,12627..12772,12871..13501))

gene

/gene="At4g03490"

CDS

complement(join(10335..10525,10983..11078,11590..12133,
 12345..12500,12627..12772,12871..13501))

/gene="At4g03490"

/note="contains similarity to mammalian ankyrins
 similarity to similar to A. thaliana hypothetical protein
 F21B7.8, GenBank accession number AC002560"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB77834.1"

/db_xref="GI:7270672"

/translation="MGKVQEFKEKWEENETPVLQOVTFQGNTHLHAAIYGHDLHVR
 ILAYELNLRNKRGLNCFVSPESHYQTLVLRNRYKGDALHVAAGHKLIVGLLI
 DCLRLQPDITMIVIGSEQMVGNIERYNSNDNTALHLSLGNHVSLSQILVREDST
 CFLDKEDVSPLYMAAAGAYVSLVEHMLRGDASFGKSVLCAAVKQNLIDILFVLE
 SDSDLVESRDEGRTPLATAASIGYDIGVOHMLTFASSTQGOVNLHVAKSGNARAV
 GYLKSDVKRLINEQDEGNTPLHLASSNSHPKVMILIMWALVAAGTTPRVRHLRAD
 IPLGTTDEDLILKHQRVNLVVAIVLVAATMAFAAGLSVPLGYNSTFEKSNVKSHE
 ESAPHAFVIGHSIAVIYTAIVSTVALIGTQLADLKMLTFKFIYVPLLGFSIIAMSLE
 VAGLYVLGHHSYWLAIIFYASGGFYLMALLLIIFYASPTFTLSRSLNSLVQMSKE
 DVDSVNLVPAPTEELAKIAIQVNRCTQGNIGVTSLSVYLLKHVAPINVLILLKLWN
 HYSLVSGHLYDGPKKTVSOL"

complement(10335..10525)

/gene="At4g03490"

/number=1

complement(10526..10982)

/gene="At4g03490"

/number=1

complement(10983..11078)

/gene="At4g03490"

/number=2

complement(11079..11589)

/gene="At4g03490"

/number=2

complement(11590..12133)

/gene="At4g03490"

/number=3

complement(12134..12344)

/gene="At4g03490"

/number=3

complement(12345..12500)

/gene="At4g03490"

/number=4

complement(12501..12626)

/gene="At4g03490"

/number=4

complement(12627..12772)

/gene="At4g03490"

/number=5

complement(12773..12870)

/gene="At4g03490"

/number=5

complement(12871..13501)

/gene="At4g03490"

/number=6

13514..15185

/note="marker mi233"

14170..17288

/gene="At4g03500"

Join(14170..14204,15103..15848,15939..16477,16650..17288)

/gene="At4g03500"

/note="hypothetical protein with some similarity to
 ankyrin
 similarity to"

/codon_start=1

/product="hypothetical protein"

/number=5

misc_feature

gene

CDS

```
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/translation="MKNESRTESEIKLIFSGVFSMGPRDPVHLDRRVAENHQAIPIR
NNVRSFSLDLSTLFDSTSEKPMDPKTMARAGKENYLRNNYSLSVAPTLVNDRG
DIICHLNASSGHVSLVRIIIOKCPGLLLKSNMGVEVALHIAAAGHLDVVMNIDFIN
NTSCNLFPAKKNQDTALHVALKKGHEVAASYLSAAKSLSFVNRDGFSP
LYLAIEAGHTSLVTVMGTSNELSSKVGGRSIVHAALKANRDKIDILLALSKDASLNL
RDEGRTSLSCFASGYGQFSYLPDKNRDVIYVSDDGLPFTHMAKYGHVQILEIL
KHCPEATELDRDQNLHLAAKYGLKVIKFLSCCKDKNKKLLNEQDVNGWTFPLH
LATLNMHPKVVMETWHRVLDLKNYIGFTALDVAEENIDSSIVHORUTWALLINA
GAPKSSPTENLRSKPDGKYKORVNTMLVATLVATVMTAGTFFGGYNDSEFP
HLGMAVAKRTAFQVFLVCDTLAMYSIIITVALIWAQGLDLSILKAFNALPFLGL
ALTSMSIAFMAGTVAYSHLPILGIVGLIGIIFLLVLLVLLVYVSPYAHQAQPLLHR
IFYYPYFLKLLAAGDNKNIVDYTASDE"
14170..14204
/gene="AT4g03500"
/number=1
14205..15102
/gene="AT4g03500"
/number=1
15103..15848
/gene="AT4g03500"
/number=2
15849..15938
/gene="AT4g03500"
```

exon

intron

exon

intron

Query Match 29.28; Score 497; DB 8; Length 159629;

Best Local Similarity 92.58; Pred. No. 5.3e-77;

Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

QY 1 TGGGTTTTTATGGATAACATGCAAAATATTTATTTATTTTCATGAGTTTTTATGGATAG 60

Db 29758 TGAGTTTTTATGGATAACATGCAAAATATTTATTTATTTTCATGAGTTTTTATGGATAG 29699

QY 61 CATGCAAAATATTTATTTATTTTCATGAGTTTTTATTTTCATGAGTTTTTATGGATAG 120

Db 29698 CATGCAAAATATTTATTTATTTTCATGAGTTTTTATTTTCATGAGTTTTTATGGATAG 29639

QY 121 TTAATCAGACATTTGTTTTTAAATCAATCTATCTATATCAACAGCATTTGAC 180

Db 29638 TTAATCAGACATTTGTTTTTAAATCAATCTATCTATATCAACAGCATTTGAC 29579

QY 181 GGAATTCAGGTAAGAGAGAAATTAAGATGAGAGATAGAGATTTCTATGGAAA 240

Db 29578 GGAATTCAGGTAAGAGAGAAATTAAGATGAGAGATAGAGATTTCTATGGAAA 29519

QY 241 AGAAGAGAGACATGTAGGTGTAACAAATAAGAGATATCATGATATTTATGAGAG 300

Db 29518 AGAAGAGAGACATGTAGGTGTAACAAATAAGAGATATCATGATATTTATGAGAG 29459

QY 301 GTGGTGAAGATTTATTTAGGAGGGGAGAGAAATAGAAAAGAAATGACATGGTAA 360

Db 29458 GTGGTGAAGATTTATTTAGGAGGGGAGAGAAATAGAAAAGAAATGACATGGTAA 29400

QY 361 TCTGAAGATGATTTGTTTAAAGATGAAGAGAGAAAGAAAGAAATGCTTAAAGTC 420

Db 29399 TTTGAAGATGATTTGTTTAAAGATGAAGAGAGAAAGAAAGAAATGCTTAAAGTC 29340

QY 421 TCGTAAAGATGATTTGTTTAAAGATGAAGAGAGAAAGAAAGAAATGCTTAAAGTC 480

Db 29339 TCGTAAAGATGATTTGTTTAAAGATGAAGAGAGAAAGAAAGAAATGCTTAAAGTC 29283

QY 481 CTAACTATGCGCAAAATTTCTGAGCGCAAAATCTATTTTGGTCCAAAGTTATTTGTG 540

Db 29282 CTAACTATGCGCAAAATTTCTGAGCGCAAAATCTATTTTGGTCCAAAGTTATTTGTG 29223

QY 541 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATCTCTTAAATAATA 589

Db 29222 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATCTCTTAAATAATA 29174

RESULT 11

ATH250185

```
LOCUS ATH250185
DEFINITION Arabidopsis thaliana mRNA for NIMIN-2 protein (nimin-2 gene).
ACCESSION AJ250185
VERSION AJ250185.1 GI:12057155
KEYWORDS nimin-2 gene; NIMIN-2 protein.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 369)
AUTHORS Weigel, R., Bauscher, C., Pfitzner, A.J. and Pfitzner, U.M.
TITLE NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins
from Arabidopsis that interact with NPR1/NIM1, a key regulator of
systemic acquired resistance in plants
JOURNAL Plant Mol. Biol. 46 (2), 143-160 (2001)
MEDLINE 2135044
PUBMED 11442055
REFERENCE 2 (bases 1 to 369)
AUTHORS Pfitzner, U.M.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Pfitzner U.M., General Virology, Institute
of Genetics, Emil Wolff Str. 14, 70599 Stuttgart, GERMANY
COMMENT Related sequence: AB023041.
FEATURES
source
1..369
/organism="Arabidopsis thaliana"
/cultivar="Col-0"
/db_xref="taxon:3702"
/tissue_type="green vegetative tissue"
/dev_stage="3-wk-old"
1..369
/gene="nimin-2"
1..369
/gene="nimin-2"
/function="putative role in SAR"
/codon_start=1
/evidence="experimental"
/product="NIMIN-2 protein"
/protein_id="CAC19845.1"
/db_xref="GI:12057156"
/translation="MNNLSKKEERVEEDNGKSDGNRGKPSPEVTVTTEBVDPEFKI
LRRVHVATRTVAKVNGVYAEGLPSKPKRSCNGLRNSLDCNGVRDGEFDEINRGL
QELGLDLACKPEPDSVLSL"
BASE COUNT 113 a 53 c 121 g 82 t
ORIGIN
Query Match 21.7%; Score 369; DB 8; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.5e-54;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 863 ATGACAACCTCTTTGAAGAAAGAACCGGTAGAGAAGATAACCGAAATCTCAGCGT 922
Db 1 ATGACAACCTCTTTGAAGAAAGAACCGGTAGAGAAGATAACCGAAATCTCAGCGT 60
QY 923 AACAGAGGAAACCGTCGACGAAAGTTGTCGACGGTAACGAGAGAGAGATGAG 982
Db 61 ACACAGGGAACCGTCGACGAAAGTTGTCGACGGTAACGAGAGAGATGAG 120
QY 983 TTTTCAAGATATTACGGAGAGTACACGTCGCGACACGACGCGTTCGCGAAAGTTAACGCG 1042
Db 121 TTTTCAAGATATTACGGAGAGTACACGTCGCGACACGACGCGTTCGCGAAAGTTAACGCG 180
QY 1043 GGTGTTGCTGAGGAGAGTTACCGTCTAAGAGAGAGAAACGAGATTCAGAAATCTCGGTTG 1102
Db 181 GGTGTTGCTGAGGAGAGTTACCGTCTAAGAGAGAGAAACGAGATTCAGAAATCTCGGTTG 240
QY 1103 AGAAACTCGTTGGATTGTAACGCGCTTCGACGCGAGAAATTCGATTAATCGGTC 1162
Db 241 AGAAACTCGTTGGATTGTAACGCGCTTCGACGCGAGAAATTCGATTAATCGGTC 300
QY 1163 GGGTACAGGGTTTGGGTTTGGATCTGAACTGTAAACCGGACACAGAGCGTTAGTTTA 1222
```

Db 301 GGGTTACAGGTTGGTTGAGTGAACGTAAACCGGAACAGACAGCGTTAGTTTA 360
 QY 1223 TCGTTGTAG 1231
 Db 361 TCGTTGTAG 369
 RESULT 12
 AC007203/c
 LOCUS
 DEFINITION Arabidopsis thaliana chromosome I BAC T10P12 genomic sequence,
 complete sequence.
 AC007203
 VERSION AC007203.3 GI:4996903
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Unpublished
 2 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (03-APR-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 3 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (04-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 4 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On Jun 4, 1999 this sequence version replaced gi:4678187.
 Bases 1-6718 of clone T10P12 overlap with bases 122,554-129,271 of
 'IGF' BAC clone F1121, gb|AC005687
 e-mail for correspondence: arab@sequence.stanford.edu
 Genes with similarity to proteins in the databases are described as
 'putative', '-like' or 'similar to'. Genes that have EST
 similarity but no significant protein similarity are described as
 'unknown proteins'. Genes that are annotated based only on gene
 prediction software are described as 'hypothetical proteins'.
 The software programs used to predict genes include: Graal
 (Informatics Group, Oak Ridge National Laboratory,
 http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
 & A. Salamov, Sanger Centre, http://genomic.sanger.ac.uk/),
 NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/NetPlantGene.html) and
 eMotif (Newill-Manning, C.G. Wu, T.D. & Brutlag, D.L.,
 http://motif.stanford.edu/projects.html).
 FEATURES
 Location/Qualifiers
 1..95190
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="I"
 /clone="T10P12"
 1..6718
 /note="overlap with bases 122,554-129,271 of 'IGF' BAC
 clone F1121, gb|AC005687"
 9474..10628
 complement(19003..23764)
 /gene="T10P12.1"
 complement(19003..23764)
 21074..21678,21933..22099,22247..22640,22931..23131,
 23362..23764)
 /gene="T10P12.1"
 /note="Similar to reverse transcriptase: Similar to reverse
 transcriptase, part of putative retrotransposon"
 /codon_start=1
 /protein_id="AAD39270.1"
 /db_xref="GI:5080761"
 /translation="MDMIFINVGVKVRKIENSKTAACAWGTDLTLLYLKSLPNRYVL
 QUKVNYRMQNSTLEENVEQFQMSDLNNIQVDPDEVAQVAILLSLPDSCDM
 LKETLYRGREGIKLDVISAALSKELELDSESTSEVNGESKGDHGEWIIDTECSFHM
 TPKEVHIDFEAKSGKVRMPNPSFLEVGIGKVRFTNQDGTSCFGNDISIKIKCFENF
 VVGKTPRTSFGTAQHVTKEDLYVHSDLWSSNPVSI SKQYFTSFDDMSRKWVIN
 PLSAVENKIPHEELSVIPSLGRLRFGCIVYVSHQECKLDPKAKKGVFVGPESHVD
 FTSSEENQRTQIARDPRRQVILPSRLKNDYEDDEVLDEIAGYAVLITEDGNGPPAS
 FOEALQSDSGKWLAADEIQSLINKTVLERNKSKQPIGCKWFKRAGIAGVE
 KPRKARLVTKGYSOKEGIYFOEIFSPVVKVYSIRLLSIVHDLMEQLQMDKFTAFLL
 HGYLEMDIEMDQEGVHEKYPKRYIYLLYVDDILIAASKKKHVKCEKLVLLNSFEM
 KOLGDRAKILGMEIINDREGGLTISOEYLILKLPATVEEVQKQKVMRTVTPYSAV
 GSLMYSGICTRPDLAHSVLCMWRKPLKLDHWQAVKWLRCIVGTMDRKLCYKNEGG
 LALEGYCSYAAADKEKRSTGVFTFGNTISWRSSI QKVVALUSKAEYTMALITDA
 TNFTVMKGLVNELGFTQSVNIHCDQSQVIALAKNAVYHRTKHIDVYHFIRELAK
 ANNELTKLEGGSLSEEEKQVLAWKKEYDDAATYNTKLVVELKELKIPPVFPNLLIP
 NSLGNRSAGELADGVGTDHASSNLRDEDMCFSEEPAP"
 29564..30696
 complement(31637..35319)
 /gene="T10P12.2"
 complement(Join(31637..32634,33345..33502,33679..33916,
 34104..34307,34383..34539,34828..35319))
 /gene="T10P12.2"
 /note="Similar to reverse-transcriptases; protein is
 flanked by LTRs, potential retrotransposon"
 /codon_start=1
 /protein_id="AAD39271.1"
 /db_xref="GI:5080761"
 /translation="MNEQVQKRIQGTGYFSGGTSPDEADSWRSRVERNFSSKCSVEYLTQ
 DIAVHFLEGDHLMWRSVTKARROADWADFAEFNAKYCPQEALEDMARFLELTQ
 GERSVREYDFEENLLVYAGRMEDDOAMRLGLRDPDRVRCRVSYATKAALVE
 TAAELMVAVVQPAVQAGVQVQOQVQOQLAHGAAPQGYTTRIGSTSNRAITGL
 LAHEVCVGLTLLVDGVHVLDFGSGVCHITPKRSASGRNIRGDPGQFSYKVGQGF
 LEVLGRKAVVEAKMIEKCEAYLMTISMPESGVQVANSIRKVOEFEDVFQSLQGLP
 PWSDDPTTIVLEPFTALLSKHPTQLQQRWIDELDLQLRGATCFSKIDLTSGYHOIPI
 AEADIRKTAFTRTYGHFWIIPFARELDALWIVDRLTKSAHLAIIRKTDGAVALIT
 KKVSVETKLHGVPVSIIVSDRSKFTSAFRAFOAEMGTQVOMSTAYHPQTDQSERT
 IQTLEMDLRMCVLDMGHWADQLSLLEPAYNNYQASIGMTPFVLYGRPCQTCLMT
 QGERSIYGANYVQETTERIGVLLKNNKVEQVQRHRSYADKRRRLEFEFEFGRVYLKMA
 MLRPNRSISETKSPYRPFVRFVVFVAYRLELTDVMAFHKVRFVFSMLRKLH
 KDDEVLAKIPEDLQPNLTLEARPGVRLERRIKELRRKKIPLIKVLWDCDCGVTEETWEP
 EVMKARF"
 35970..37136
 37249..42330
 /gene="T10P12.3"
 complement(Join(37249..37347,37714..38098,38265..38532,38897..39097,
 39215..39508,39603..40889,41069..42330))
 /gene="T10P12.3"
 repeat_region
 gene
 CDS
 repeat_region
 gene
 CDS


```

QY 125 AATCAGACATTGTTTAAATCAAACTAATCTCTTATATCAACAGCAGCATTCGACGAA 184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1473 ATCTATGCAACTAGTTAAAGATAGAGAAATTTGGTACRRRRRRRRRRRRRRRRR 1414
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 AATTGAGTAAAGAGAAAATAAGAAATGAGATAGAGATTTCTATGGAAGAA 244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1413 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 AGAGACACATGTAGGTGACAAATAAAGAGATATGATGATATATTTATGAGAGTGG 304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1353 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 305 TGAAGATTATTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATGCTGAACTG 364
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1293 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1234
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 365 AAGAAGATGAATTTGTTTAAAGATGAAGAGAGAAGCAATCCTGCTAAATCTCGT 424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1233 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 425 AAAGAAGATGAAGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1173 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14

AC115878/c

LOCUS

DEFINITION AC115878 171763 bp DNA linear HTG 14-JUN-2002
Mus musculus clone RP24-358G9, WORKING DRAFT SEQUENCE, 16 ordered
pieces.

AC115878

AC115878.2 GI:21426394

HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 171763)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-358G9

Unpublished

2 (bases 1 to 171763)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhgaltier, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, K., LaRocque, K., Lamazares, R.,
Kamat, A., Karatas, A., Kells, C., LeRocque, K., Lindblad-Toh, K., Liu, G.,
Linders, T., Lehoczy, J., Levine, R., Riley, R., Rise, C., Rogov, P.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
O'Neil, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 171763)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhgaltier, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mienga, V., Mienga, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schuback, R., Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 14, 2002 this sequence version replaced gi:19683557.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L24766

Center clone name: 358_G_9

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 167247 bases at least Q40

Consensus quality: 168964 bases at least Q30

Consensus quality: 169725 bases at least Q20

Insert size: 170000; agarose-fp

Quality coverage: 6.3 in Q20 bases; agarose-fp

Quality coverage: 6.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 1159: contig of 1159 bp in length

* 1160 1259: gap of 100 bp

* 1260 2312: contig of 1053 bp in length

* 2313 2412: gap of 100 bp

* 2413 3453: contig of 1041 bp in length

* 3454 3553: gap of 100 bp

* 3554 5119: contig of 1566 bp in length

* 5120 5219: gap of 100 bp

* 5220 7819: contig of 2600 bp in length

* 7820 7919: gap of 100 bp

* 7920 9949: contig of 2030 bp in length

* 9950 10049: gap of 100 bp

* 10050 14159: contig of 4110 bp in length

* 14160 14259: gap of 100 bp

* 14260 18550: contig of 4291 bp in length

* 18551 18650: gap of 100 bp

* 18651 25307: contig of 6657 bp in length

* 25308 25407: gap of 100 bp

* 25408 34087: contig of 8680 bp in length

* 34088 34187: gap of 100 bp

*	34188	50517:	contig of 16330 bp	in length
*	50518	50617:	gap of 100 bp	
*	50618	65071:	contig of 14454 bp	in length
*	65072	65171:	gap of 100 bp	
*	65172	81034:	contig of 15883 bp	in length
*	81035	81154:	gap of 100 bp	
*	81155	95034:	contig of 13930 bp	in length
*	95085	95184:	gap of 100 bp	
*	95185	133137:	contig of 38013 bp	in length
*	133198	133297:	gap of 100 bp	
*	133298	171763:	contig of 38466 bp	in length.

FEATURES

```

location=qualifiers
source
1. .171763
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-358G9"
/clone_lib="RPC1-24 Male Mouse BAC"
1. .1159
/note="assembly_fragment"
1260. .2312
/note="assembly_fragment"
2413. .3453
/note="assembly_fragment"
3554. .5119
/note="assembly_fragment"
5220. .7819
/note="assembly_fragment"
7920. .9949
/note="assembly_fragment"
10050. .14159
/note="assembly_fragment"
14260. .18550
/note="assembly_fragment"
18651. .25307
/note="assembly_fragment"
25408. .34087
/note="assembly_fragment"
34188. .50517
/note="assembly_fragment"
50618. .65071
/note="assembly_fragment"
65172. .81054
/note="assembly_fragment"
81155. .95084
/note="assembly_fragment"
95185. .133197
/note="assembly_fragment"
133298. .171763
/note="assembly_fragment"
BASE COUNT 48403 a 34526 c 33542 g 53768 t 1524 others
ORIGIN

```

Query Match 5.3%; Score 90.2; DB 2; Length 171763;
Best Local Similarity 56.1%; Pred. No. 1.6e-06;
Matches 170; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

[illegible][illegible]

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0293L09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primers; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 277729 bases at least Q40
Consensus quality: 279043 bases at least Q30
Consensus quality: 279926 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 181848; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 9.94 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 8161: contig of 8161 bp in length
* 8162 8261: gap of unknown length
* 8262 51021: contig of 42760 bp in length
* 51022 51121: gap of unknown length
* 51122 84880: contig of 33759 bp in length
* 84881 84980: gap of unknown length
* 84981 280982: contig of 196002 bp in length.
* Location/Qualifiers

```

FEATURES

```
source
1. .280982
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-299L9"
1. .8161
misc_feature
/feature="assembly_name:Contig10"
8262..51021
misc_feature
/feature="assembly_name:Contig11"
51122..84880
misc_feature
/feature="assembly_name:Contig12"
84981..280982
misc_feature
/feature="assembly_name:Contig13"
90913 a 51036 c 51472 g 87256 t 305 others
BASE COUNT
ORIGIN
Query Match 5.2%; Score 89.2; DB 2; Length 280982;
Best Local Similarity 54.6%; Pred. No. 2.2e-06;
Matches 178; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 138 TTTTAAATCAATCTATCTCTTATATACACACGACATTGACGGAAATTCAGGTAAA 197
Db 162899 TTTTAACTGAAAAAGCAAACTCATCAAAAAGAAAGAAAGCAAGAAAGAGAGAGAGA 162840
QY 198 AGAGAAAAATAAGATGAGAGATAGAGAGATTCTTATGCAAAAAGAAAGAGAACATGT 257
Db 162839 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162780
QY 258 AGGTGAACAAATAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTTT 317
Db 162779 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162720
QY 318 AGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 377
Db 162719 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162660
QY 378 GTGTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 437
Db 162659 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162600
QY 438 AAGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
Db 162599 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162574
```

Search completed: November 29, 2002, 00:40:31
Job time : 5728 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 22:59:21 ; Search time 2040 seconds
(without alignments)
13496.239 Million cell updates/sec

Title: US-09-733-685-3

Perfect score: 1700

Sequence: 1 tgggttttttggataaca.....tttcagacttacgaacatat 1700

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estm:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	13.5	230	17	BH866477
2	203.8	12.0	531	17	BH696479
3	201.2	11.8	206	17	BH866474
4	185	10.9	237	17	BH814587
5	175.2	10.3	196	17	BH866500
6	173.6	10.2	267	17	BH854168

C	7	128.4	7.6	387	17	BH749275
C	8	101.6	6.0	1101	17	CNS0021J
C	9	96	5.6	660	17	BH183498
C	10	96	5.6	660	17	CNS070NJ
C	11	91	5.4	982	17	AQ325799
C	12	89	5.2	529	9	AL514657
C	13	88	5.2	768	17	CNS01V5E
C	14	86.6	5.1	1101	17	CNS008VL
C	15	86.4	5.1	764	17	AZ193158
C	16	85.8	5.0	842	17	CNS010QN
C	17	85.6	5.0	997	17	CNS005TE
C	18	85	5.0	886	17	CNS075DF
C	19	85	5.0	934	17	AZ184244
C	20	85	5.0	1201	17	CNS0167M
C	21	84.8	5.0	1101	17	CNS001T2
C	22	84	4.9	647	17	BH038922
C	23	84	4.9	839	14	BQ151187
C	24	83.8	4.9	922	17	AZ548363
C	25	83.4	4.9	638	17	AQ329262
C	26	83.4	4.9	848	17	AZ549808
C	27	82.6	4.9	89	17	BH866492
C	28	82.6	4.9	718	17	AG160666
C	29	82.4	4.8	434	17	CNS060Z3
C	30	82.4	4.8	952	17	AQ897460
C	31	82.2	4.8	580	17	CNS008HN
C	32	82.2	4.8	1101	17	CNS0039G
C	33	82	4.8	888	17	AZ549422
C	34	81.8	4.8	625	17	CNS036A2
C	35	81.8	4.8	815	17	AG044049
C	36	81.4	4.8	891	17	CNS009JU
C	37	81.2	4.8	1001	17	CNS0064B
C	38	81.2	4.8	1084	17	CNS06PLB
C	39	81.2	4.8	1101	17	CNS012TP
C	40	81	4.8	89	17	BH866501
C	41	80.8	4.8	1042	17	CNS0148K
C	42	80.6	4.7	797	10	BE034282
C	43	80.6	4.7	1101	17	CNS0181N
C	44	80.4	4.7	614	17	CNS0152H
C	45	79.8	4.7	987	17	CNS00418

ALIGNMENTS

RESULT 1	BH866477	BH866477	230 bp	DNA	linear	GSS 05-AUG-2002
LOCUS	BH866477	SALK_101380	Arabidopsis thaliana	TDNA	insertion lines	Arabidopsis
DEFINITION	BH866477	thaliana genomic clone SALK_101380	DNA	sequence.		
ACCESSION	BH866477					
VERSION	BH866477.1	GI:22102375				
KEYWORDS	GSS.					
SOURCE	thale cress.					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.					
AUTHORS	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.					
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome					
JOURNAL	Unpublished (2001)					
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA.					


```

58 a 31 c 72 q 45 t
BASE COUNT
/clone.lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic site at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html."

```

Query Match	11.8s;	Score	201.2;	DB	17;	Length	206;
Best Local Similarity	98.5s;	Pred.	No. 6.8e-24;				
Matches	203;	Conservative	0;	Mismatches	3;	Indels	0;
Gaps	0;						
Qy	938	TCGACGGAAGTTGTTCCGACGGTAA	CGGAGGAAGCTGGATCAGATTTTCAAGATATTA	997			
Db	1	TCGACGACCTTGATCGACGGTAC	GGGAGAGGTTGGATGAGTTTTCAAGATATTA	60			
Qy	998	CGGAGAGTACAGCTGGCGACAC	AGACGGTTGCGAAAGTTAACCGCGGTGTGCTGAGGGA	1057			
Db	61	CGGAGAGTACAGCTGGCGACAC	AGACGGTTGCGAAAGTTAACCGCGGTGTGCTGAGGGA	120			
Qy	1058	GAGTTACCGCTCTAAGAACAGGAA	ACGGAGTCAGAACTCTGGGTTGAGAACTCGTTGGAT	1117			
Db	121	GAGTTACCGCTCTAAGAACAGG	AAACCGAGTCAGAACTCTGGGTTGAGAACTCGTTGGAT	180			
Qy	1118	TGTAACGGCGGTTTCGAGACGG	GAGAATT	1143			
Db	181	TGTAACGGCGGTTTCGAGACGG	GAGAATT	206			

RESULT 4	
BH814587	
LOCUS	
DEFINITION	237 bp DNA linear
ACCESSION	SALK_066674 Arabidopsis thaliana TDNA insertion lines
VERSION	thaliana genomic clone SALK_066674, DNA sequence.
KEYWORDS	BH814587
SOURCE	BH814587.1 GI:20394318
	GSS.
	thale cress.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustersids II; Brassicales; Brassicaceae; Arabidopsida;
1 (bases 1 to 237)

REFERENCE

AUTHORS

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk R., Gadrinab
, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Ecker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salic.edu
This is single pass sequence recovered from the left border of TDNA.

```

FEATURES
source
Class: tDNA tagged.
Location/Qualifiers
1. .237
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_066674"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can

```

BASE COUNT	82 a	38 c	63 g	53 t	1 Others
ORIGIN	be found at http://signal.salk.edu/~dna_protocols.html				

Query Match	10.9%	Score 185;	DB 17;	Length 237;
Best Local Similarity	90.4%	Pred. No. 3.2e-21;		
Matches 208;	Conservative	0;	Mismatches 21;	Indels 1;
				Gaps 1;

Qy	733	ATATCGACCGTCTAAGTTCACACCGAGCGGTATAGAGTTTCATTATAATTTAGCAA	792
Db	1	ATATCGACCGTCTAAGTTCACACCGAGCGGTATAGAGTTTCATTATAATTTAGCAA	60
Qy	793	AATAAAATCAGCAATAATTTTTCTTGACTAAGCTTAAACGAGCGCGTTAAACATTTCT	852
Db	61	AATAAAATCAGCAATAATTTTTCTTGACTAAGCTTAAACGAGCGCGTTAAACATTTCT	120
Qy	853	TCTGGCTAACATGAACAACCTTTTGAAGAAGAAGACGGCTAGAAGAAGATAACGGAA	911
Db	121	TGTGCTAGGNTTGCACCACTTTTGAAGAAGAAGACGGCAAGCAAGAAGATACCGAAG	180
Qy	912	AATCTGACGGTAAACAGAGGAAACCGTCGACGGAAGTTGTCGACGGTA	961
Db	181	AATCTGACGGTAAACAGAGGAAACCGTCGACGGAAGTTGTCGACGGTA	230

RESULT 5				
BH866500				
LOCUS	BH866500	196 bp	DNA	linear
DEFINITION	SALK_101406 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_101406, DNA sequence.			

ACCESSION	BH866500	
VERSION	BH866500.1	GI:22102398
KEYWORDS	GSS.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	

REFERENCE
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab
J.C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J., and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)

CONTACT: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SUGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

```

Class: TDNA lagged.
FEATURES
    source
        Location/Qualifiers
            1. .196
                /organism="Arabidopsis thaliana"
                /strain="Columbia 0"
                /db_xref="taxon:3702"
                /clone="SALK_101406"
                /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                /note="PCR was performed on Arabidopsis thaliana lines
                    each of which contains one or more TDNA insertion
                    elements. The resultant fragment for each line was
                    directly sequenced to determine the genomic sequence at
                    the site of insertion. Details of the protocols used can
                    be found at http://signal.salk.edu/~dna\_protocols.html"
            50 a 33 c 64 g 49 t

```

BASE COUNT	50 a	33 c	64 g	49 t
ORIGIN				
Query Match	10.3%; Score 175.2; DB 17; Length 196;			
Best Local Similarity	93.4%; Pred. No. 1.5e-19;			

Db 80 TGAGTTTATTGGATAACATGACAAATATTATATATACATGAATTTTGTGGATAG 139

284 AGAGAGAACATGTGGGTGATCAAAAAATAAGAGATATGATGATATATTTTATGAGAGGTG 222

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

982 bp	DNA	linear	GSS 08-JAN-1999
nbx0021B14r	CUGI Rice BAC Library	Oryza sativa	genomic clone
nbx0021B14r,	DNA sequence.		
AQ325799			
AQ325799.1	GI:4117649		
GSS.			
Oryza sativa.			
Oryza sativa			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.			
1 (bases 1 to 982)			
Wing, R.A. and Dean, R.A.			
A BAC End Sequencing Framework to Sequence the Rice Genome			
Unpublished (1998)			
Contact: Wing RA			
Clemson University Genomics Institute			
Clemson University			
100 Jordan Hall, Clemson, SC 29634, USA			
Tel: 864 656 7288			
Fax: 864 656 4293			
Email: rwing@clemson.edu			
Seq primer: GGAAACAGCTATGCACATG			
Class: BAC ends			
High quality sequence start: 4			
High quality sequence stop: 123.			

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

1: 1902
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nxb0021B14r"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
141 a 69 c 43 q 674 t 55 others

```

[illegible]

```
Query Match          5.4%; Score 91; DB 17; Length 982;
Best Local Similarity 46.9%; Pred. No. 7.7e-06;
Matches 229; Conservative 0; Mismatches 259; Indels 0; Gaps 0;
```

QY 9 TATTGGATAACATGACAAATATTTTATTTATTTCAATGAGCTTTTATTTGGATAGCATGACAA 68
||||| - - - - - ||| - - - ||||| - - - - -
Db 903 TATGGAAATATGAAAAGAAATTTTAATTTATATAAAATAAATTTTAATTTAATTTAATAAT 844

QY 69 ATATTAAATATATCAGTGTTAAATAACATGTTTGTCTTAAAAATACATGCATTTTAAATC 128
 |||||
 Db 843 TGATTAAATATATATAAATATATATATTTTAAAAATATAATATTAATTAATTTAATAATT 784

QY 129 AGACATTTGTTTAAATCAAATCTAATCTTATATACAAACGRCATTGACGGAATTT 188

Db	783	AAATAATNTATGGATAAATAATATATAATTAATTTAATATATATAAAATAAATA	724
Qy	189	CAGGTAAAGACAGAGAAATAAAGAAATGAGAGATACAGAGATTTCTATCGAAAGAAAGAGAG	248
Db	723	AAAAAATAAAAAAANNNAAAAAATAATTAATAAAAAAANNNAAAAAANNNAAAAA	664
Qy	249	AGAACATGTAGGTGAACACAAATAAAGAGATATGATGATATATTTATCGAGAGGTGGTGAA	308
Db	663	AANAANAATAAAAAAANNNAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	604
Qy	309	GATTATTTTGGAGAGGGAGAGAGAAATAGAAAAGAAAATGCATGGTGAATCTCGAAGA	368
Db	603	AAATANATAAANAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	544
Qy	369	AGATGAATTGTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTTAAGTCTCGTAAAG	428
Db	543	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	484
Qy	429	AAAGATGAAGAAAGAAACAAAGAGGAAGAAAGAGAGAAAGAGTAAAGTCTCGTAAAG	488
Db	483	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	424
Qy	489	TGCCAAAAA 496	
Db	423	AAAAAATAAAAAA 416	

RESULT 12
AT.514657

LOCUS	AL514657	529 bp	linear	EST 13-FEB-2001
DEFINITION	AL514657 LTI_NFL006_PL2	Homo sapiens	cdna clone	Cl0BB002ZE09.3
	prime, mRNA sequence.			

ACCESSION AL514657
VERSION AL514657.1
KEYWORDS EST.
SOURCE human

BOOKS
ORGANISM
HUMAN.
HOMO SAPIENS

ORGANISM HOMO SAPIENS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 529)

AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscone

CONTACT : genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seoref@genoscope.cns.fr Web : www.genoscope.cns.fr

FEATURES

source

33 Tmoe

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBB002ZF09"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPOT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPOT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :

```

BASE COUNT	364 a	20 c	9 g	89 t	47 others
ORIGIN	http://murroughen.invallogen.com				

Query Match 5.2%; Score 89; DB 9; Length 529;

Best Local Similarity 43.9%; Pred. No. 2.1e-05;
Matches 206; Conservative 35; Mismatches 228;
Indels 0; Gaps 0;

QY 28 TATTTATTATTTTCATGAGTTTTTTATTGGTAGCATGACAAATATTANTATATCAGTCTT 87

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone.lib="RPCI-98"
/note="end : 17"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 5.1%; Score 86.6; DB 17; Length 1101;
Best Local Similarity 36.8%; Pred No. 4e-05;
Matches 215; Conservative 96; Mismatches 270; Indels 4; Gaps 2;

```
QY 5 TTTTATTGGAATCAATGACAAATATTATTATTTCATGAGTTTATTGCGATGACGATG 64
|| || || || || || || || || || || || || || || || || || || || || ||
Db 487 TTATWAATWAAAAAANWAAATTTTWWWWTTWTTTWWAAWTTWATAAAAAAAW 546

QY 65 ACAATATTAAATATACATGCTTAATACATGTTTGTTCITTAATAATACATGATTTAA 124
|| || || || || || || || || || || || || || || || || || || || || ||
Db 547 ATAAATTAANWAAATWATTAATAATTTTAAWAAWTTATATTAATWTTAAATTTWTTAA 606

QY 125 AATCAGACATTTGTTTAAATCAATCTCTATATACACAGCATGTGACGGAA 184
|| || || || || || || || || || || || || || || || || || || || || ||
Db 607 ATAAAAAATATTTTWTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 666

QY 185 AATTCAGGTAAAGAGAAAAATTAAGATGAGAGATGAGAGATTTCTATGCGAAAAAGAA 244
|| || || || || || || || || || || || || || || || || || || || || ||
Db 667 WWTAAATTAATAATTTWAAATAWAAAAAATAAAAAAATAWAAWAAATWATAWATAWT 726

QY 245 AGAGAAACATGTAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAGGTGG 304
|| || || || || || || || || || || || || || || || || || || || || ||
Db 727 TAAAWAAATAAAWAAWAAATWAAATATATWATATATATWTTWAAWTTWAAWTTWAAWTTAT 786

QY 305 TGAAGATTATTTAGAGAGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTG 364
|| || || || || || || || || || || || || || || || || || || || || ||
Db 787 AATWAAATWAAWAAWAAATAATWATAWATAWAAWAAWAAWAAWATAWATAWATAWATAW 846

QY 365 AAGAAGATGAATTTGTTAAAGATGAAGAGAGAAAGAAACCCATGGCTAAAGTCTCGT 424
|| || || || || || || || || || || || || || || || || || || || || ||
Db 847 WAAAAAATWTAATATWATAWATAW---AAAAATAWAAWTTTWTWTTTWWAAWATAT 903

QY 425 AAGAAGATGAANAAGAAACAAAGAGGAAGAAAGAGAGAAAGCGTAAATAGACATAA 484
|| || || || || || || || || || || || || || || || || || || || || ||
Db 904 AAAWAAWATAWAAAAAATAAAAAATAAAWAAWTTWATTTTATTTAAATWTTATWAA 963

QY 485 CTATTCGCAAAATTTCTGTAGCGCAAAATACTATTGTGTCACAGTTATTTTGTGTTAT 544
|| || || || || || || || || || || || || || || || || || || || || ||
Db 964 TTWTTWAAWTTWATATWTTW-TATWATATATWTTAAWATAATATTTTATTAWAAWTTAT 1022

QY 545 CTTTGAAGTCAAAAGTTATTTCTTACATATACTCTAAAAATATA 589
|| || || || || || || || || || || || || || || || || || || || || ||
Db 1023 TTTAAWAAWTTAATATATWAAWTTWATAWATAWAAWAAATATWTTA 1067
```

RESULT 15
AZ193158/c
LOCUS
DEFINITION
AZ193158
VERSION
KEYWORDS

AZ193158 764 bp DNA linear GSS 30-AUG-2000
SP_1022_B2_E10_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate:1022 Col=20 Row=J, DNA sequence.
AZ193158
AZ193158.1 GI:8376337
GSS.

SOURCE
ORGANISM

Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE
AUTHORS

1 (bases 1 to 764)
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.

TITLE

A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)

JOURNAL
MEDLINE

20402566

COMMENT

Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1022 row: J column: 20
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 764.

FEATURES

source

1..764
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone.lib="Plate:1022 Col=20 Row=J"
/clone.lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 173 a 56 c 48 g 469 t 18 others
ORIGIN

Query Match 5.1%; Score 86.4; DB 17; Length 764;
Best Local Similarity 47.6%; Pred. No. 5e-05;
Matches 234; Conservative 0; Mismatches 258; Indels. 0; Gaps 0;

```
QY 5 TTTTATTGATAACATGACAAATATTATTATTTCATGAGTTTATTTCGATAGCATG 64
|| || || || || || || || || || || || || || || || || || || || || ||
Db 728 TATTATAATATAGATATATAGATTTATTATATTATAGATAATANTTAGAAGAATT 669

QY 65 ACAATATTAAATATACAGTGTTAATAACATGTTTGTCTTAAATACATGCAATTTAA 124
|| || || || || || || || || || || || || || || || || || || || || ||
Db 668 AATAATATTTTAAATAAAAAATAATATATATAAAAAATAAAAAATAATATTTTATAA 609

QY 125 AATCAGACATTTGTTTAAATCAAATCTAATCTCTTATATCACACGACATTCACGGA 184
|| || || || || || || || || || || || || || || || || || || || || ||
Db 608 TTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549

QY 185 AATTCAGGTAAAAAGAGAAAAATAAGAAATGAGAGATGAGAGATTTCTATCGAAAAAGAA 244
|| || || || || || || || || || || || || || || || || || || || || ||
Db 548 ATATAAAATTTAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 489

QY 245 AGAGAAACATGTAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAGGTGG 304
|| || || || || || || || || || || || || || || || || || || || || ||
Db 488 TAAAAAATAAANAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 429

QY 305 TGAAGATTATTTAGGAGGGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTG 364
|| || || || || || || || || || || || || || || || || || || || || ||
Db 428 AAAAAAATAAANAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 369

QY 365 AAGAAGATGAATTTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGT 424
|| || || || || || || || || || || || || || || || || || || || || ||
Db 368 AAAAAAATAAANAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 309

QY 425 AAGAAGATGAAGAAGAAACAAAGAGAGAAAGAGAAAGCGTAAATAGACTAA 484
|| || || || || || || || || || || || || || || || || || || || || ||
Db 308 AAAAAAATAAANAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 249
```

OY 485 CTATTGCCAAAA 496
|
|
|
|
Db 248 AAAAAAAAAAAAA 237

Search completed: November 29, 2002, 00:31:28
Job time : 2058 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 23:12:21 : Search time 70 Seconds
(without alignments)
9352.804 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....tttaagacttaacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 19257720 residues
Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	73.8	4.3	529	10	US-09-983-965-2109 Sequence 2109, Ap
C 2	72	4.2	446	10	US-09-960-352-3400 Sequence 3400, Ap
C 3	71.2	4.2	152331	9	US-10-095-407-16 Sequence 16, Appl
C 4	71.2	4.2	176373	9	US-10-095-407-17 Sequence 17, Appl
C 5	68	4.0	292	10	US-09-867-701-9583 Sequence 9583, Ap
C 6	68	4.0	516	10	US-09-960-352-5785 Sequence 5785, Ap
C 7	67.8	4.0	180216	10	US-09-835-232-6 Sequence 6, Appl
C 8	66.8	3.9	174424	10	US-09-967-768A-314 Sequence 314, App
C 9	66.4	3.9	299	10	US-09-867-701-9596 Sequence 9596, Ap
C 10	66.2	3.9	307	10	US-09-867-701-9416 Sequence 9416, Ap
C 11	65.6	3.9	42999	9	US-09-799-462A-17 Sequence 17, Appl
C 12	65.6	3.9	42999	9	US-10-125-767-17 Sequence 17, Appl
C 13	65.2	3.8	451	10	US-09-960-352-10262 Sequence 10262, A
C 14	63.8	3.8	174424	10	US-09-967-768A-314 Sequence 314, App
C 15	63.4	3.7	428	10	US-09-960-352-573 Sequence 573, App
C 16	63	3.7	299	10	US-09-867-701-9477 Sequence 9477, A
C 17	63	3.7	341	10	US-09-960-352-12302 Sequence 12302, A
C 18	62.6	3.7	152331	9	US-10-095-407-16 Sequence 16, Appl
C 19	62.4	3.7	1721	9	US-09-938-842A-3578 Sequence 3578, Ap

20	62.4	3.7	4167	10	US-09-764-878-282	Sequence 282, App
21	62.4	3.7	4167	10	US-09-764-860-1145	Sequence 1145, Ap
C 22	62.4	3.7	4167	10	US-09-764-846-345	Sequence 345, App
C 23	62	3.6	425	10	US-09-960-352-4010	Sequence 4010, Ap
C 24	62	3.6	8317	10	US-09-764-869-1279	Sequence 1279, Ap
C 25	61.8	3.6	344	10	US-09-960-352-1036	Sequence 1036, Ap
C 26	61.8	3.6	431	10	US-09-960-352-5558	Sequence 5558, Ap
C 27	61.4	3.6	26555	9	US-09-860-670-161	Sequence 161, App
C 28	61.2	3.6	90541	10	US-09-759-359A-3	Sequence 3, Appli
C 29	60.4	3.6	463	10	US-09-960-352-7186	Sequence 7186, Ap
C 30	60	3.5	419	10	US-09-960-352-11234	Sequence 11234, A
C 31	59.2	3.5	393	10	US-09-960-352-4582	Sequence 4582, Ap
C 32	58.4	3.4	408	10	US-09-960-352-1221	Sequence 1221, Ap
C 33	57.8	3.4	172637	10	US-09-805-458A-3	Sequence 3, Appli
C 34	56.8	3.3	659158	9	US-09-771-208-20	Sequence 20, Appl
C 35	56.2	3.3	298	10	US-09-867-701-8978	Sequence 8978, Ap
C 36	56	3.3	42999	10	US-09-740-029-3	Sequence 3, Appli
C 37	55.8	3.3	545	10	US-09-878-374-4299	Sequence 4299, Ap
C 38	55.8	3.3	1267	12	US-10-001-843-45	Sequence 45, Appl
C 39	55.6	3.3	277	10	US-09-867-701-8971	Sequence 8971, Ap
C 40	54.8	3.2	368	10	US-09-834-975-47	Sequence 47, Appl
C 41	54.8	3.2	439	10	US-09-960-352-1009	Sequence 1009, Ap
C 42	54.6	3.2	349	10	US-09-867-701-9513	Sequence 9513, Ap
C 43	54.6	3.2	423	10	US-09-864-761-18355	Sequence 18355, A
C 44	54.6	3.2	465237	10	US-09-933-267A-1	Sequence 1, Appli
C 45	54.6	3.2	640681	10	US-09-790-988-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-983-965-2109/c
; Sequence 2109, Application US/09983965
; Patent No. US20030137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2109
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (335)
; OTHER INFORMATION:
; OTHER INFORMATION: Clone ID: 34-LIB3057-015-Q1-K1-A6
US-09-983-965-2109

Query Match 4.3%; Score 73.8; DB 10; Length 529;
Best Local Similarity 49.2%; Pred. No. 6.6e-06;
Matches 192; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY	107	AAATACATGCAATTTTAAATCAGACATTTGTTTAAATCAAACTTAATCTTTATATC	166
Db	474	AAAAAAAAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAA	415
QY	167	ACAACGACATTCGCGAAATTCAGGTAAAGAGAAATTAAGATGAGATAGAGAG	226
Db	414	AAAAAAAAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAA	355


```
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-17

Query Match          4.2%; Score 71.2; DB 9; Length 176373;
Best Local Similarity 54.6%; Pred. No. 9.9e-05;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AATAAAGATGAGATAGATAGAGATTCTTATGGAAGAAAGAGAGAACATGTAGGTG 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144774 AGACAAGAAAGAGAGAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 144715

QY 263 AACAAATAAGAGATATGATATATATTTATGAGAGGTGGTGAAGATATTTTAGGAG 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144714 AGAAAGAAAGAAAGAGAGAGAGGGAGGGAGGAGAGAGAGAGAGAGAGAGAGAG 144655

QY 323 AGGAGAGAGAGAAATAGAAAAGAAATGACATGCTGAATCTGAAGAGATGAATTCGTT 382
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144654 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 144595

QY 383 AAGATGAAGAGAGAGAGAGAGAGTCCATGGCTTAAAGTCTCGTAAAGAGATGAAAAGAA 442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144594 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144535

QY 443 ACAAAGAGAGAGAGAGAA 462
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144534 AGGAAGGAGAGAGAGAGAA 144515

RESULT 5
US-09-867-701-9583
; Sequence 9583, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9583
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9583

Query Match          4.0%; Score 68; DB 10; Length 292;
Best Local Similarity 52.9%; Pred. No. 7.6e-05;
Matches 146; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 196 AAGAGAGAAATAAGAAATGAGATAGAGAGATTTCTATGGAAGAAAGAGAGAGACAT 255
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 ACAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69

QY 256 GTAGGTGAACAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTATT 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 GAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 129

QY 316 TTAGGAGGGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTCGAAGAAGATGAA 375
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 GAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189

QY 376 TTCTGTTAAAGATGAAGAGAGAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGA 435
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 190 AAACGAGCGAGGGAGGAGAGCGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 249
QY 436 AAAGAAACAAGAGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGGC 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTC 285

RESULT 6
US-09-960-352-5785/c
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Wyatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; NAME/KEY: unsure
; LOCATION: (76),(90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785
```

```
Query Match          4.0%; Score 68; DB 10; Length 516;
Best Local Similarity 48.2%; Pred. No. 8.8e-05;
Matches 217; Conservative 0; Mismatches 232; Indels 1; Gaps 1;

QY 68 AATATTATATATACATGCTTAATACATGTTTGTCTTAAATACATGCAATTTTAAAT 127
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 AAAAAAATAATTAATATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 442

QY 128 CAGCATTTGTTTTAAATCAAACTAATCTCTTATATCAACAAGACATTCACGGAAT 187
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 441 ATAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 382

QY 188 TCAGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 247
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 381 TAAAAATAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 322

QY 248 GAGACATGTAGTGAACAAAAATAAAGAGATATGATATATATTTTATGAGAGGTGTA 307
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 262

QY 308 AGATTATTTTAGGAGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 AAAATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 202

QY 368 AAGATCAATTCGTTAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 AAAAT-AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 143

QY 428 GAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 83

QY 488 TTGCCAAAATTTCTGTAGCCGACAAATACT 517
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82 ACCCTCNATACCTTATTGCCATTGAATTT 53
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 7
US-09-835-232-6
; Sequence 6, Application US/09835232
```


GENERAL INFORMATION:
APPLICANT: Haglaczkv, Gyula
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: Heller Ehrman White & McAuliffe LLP
STREET: 4350 La Jolla Village Drive, 7th Floor
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/125,767
FILING DATE: 17-Apr-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/724,693
FILING DATE: 28-NOV-2000
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-APR-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-AUG-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-JUL-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-402J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5360
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17
Query Match 3.9%; Score 65.6; DB 9; Length 42999;
Best Local Similarity 52.5%; Pred. No. 0.00084;
Matches 166; Conservative 0; Mismatches 149; Indels 1; Gaps 1;
Qy 166 CACACACATGCGGAAATTCAGTAAAGAGAGAAATAGAAATGAGAGATAGAGA 225
Db 27215 CAGACAGACGGGGAGGAGAGACAAACAGAGAGAGAGAGAGAGAGAGAGAGA 27156
Qy 226 GATTCTATGAAAAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGA 285
Db 27155 GACAGACAGACAGACAGACAGACAAAGAACAGACAGAGAGAGAGAGACGGA 27096
Qy 286 TATATTTTATCAGAGGTGGTGAAGATTATTTAGGAGAGGGAGAGAGAAATAGAAAAGA 345
Db 27095 GAGAGAGTGTAGTGGAGGGAGAGAGACATGGAGGGAGAGACAGACAGAGAGAGA 27036
Qy 346 AAATGACATGGTGNATCTGAAGAGATGAATTCTGTAAAGATGAACAGAGAAAGAGAAC 405
Db 27035 AACAGACAGAAAGAGAGAGACGAGAGAGAGTGTGAGAGAGAGAGAGAGAGACA 26976

Qy 406 TCATGCTAAAGTCTCGTAAAGA-AGATGAAAAAGAAACAAAGAGAGAGAGAGAGA 464
Db 26975 TGGAGGAGAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 26916
Qy 465 GAAAGGCTAAATAGA 480
Db 26915 GTGAGTGTGAGAGAGAGA 26900
RESULT 13
US-09-960-352-10262/c
; Sequence 10262, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10262
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 44-LIB3058-031-Q1-K1-C8
US-09-960-352-10262
Query Match 3.8%; Score 65.2; DB 10; Length 451;
Best Local Similarity 51.7%; Pred. No. 0.0003;
Matches 148; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
Qy 183 AAATTCAGTTAAAGAGAGAAATTAAGCATGAGAGATAGAGAGATTTCTATCGAAAAAG 242
Db 438 AAAAGAGAAAAAAGAAAAAATAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 379
Qy 243 AAAGAGAGACATGTAGGTGAACAAATAAAGAGATATGATGATATATTTATGAGAGGT 302
Db 378 AAAAAGAAAAAATAAAGAAAAAATAAAGAAAAAATAAAGAAAAAATAAAGAAAAA 319
Qy 303 GGTGAAGATTATTTAGGAGGGAGAGAGAAATAGAAAAAGAAATGACATGCTGAATC 362
Db 318 AAATAAAAAG 259
Qy 363 TGAAGAAGATGAATTGTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
Db 258 AAAAGAGAAAAATTAAG 199
Qy 423 GTAAAGAGAGATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
Db 198 AAATACAG 153
RESULT 14
US-09-967-768A-314/c
; Sequence 314, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111

